EVALUATING A NOVEL EYE TRACKING TOOL TO DETECT INVALID RESPONDING IN NEUROCOGNITIVE ASSESSMENT

by

David M. Barry

Dissertation submitted to the Faculty of the
Medical and Clinical Psychology Graduate Program
Uniformed Services University of the Health Sciences
In partial fulfillment of the requirements for the degree of
Doctor of Philosophy 2014

UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814

FINAL EXAMINATION/PRIVATE DEFENSE FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE MEDICAL AND CLINICAL PSYCHOLOGY GRADUATE PROGRAM

Name of Student: David Barry

Date of Examination: May 7, 2014

Time: 8:00am

Place: B4004

DECISION OF EXAMINATION COMMITTEE MEMBERS:

	PASS	FAIL
Dr. Andrew Waters DEPARTMENT OF MEDICAL AND CLINICAL PSYCHOLOGY Committee Chairperson	✓	
Dr. Mark Ettenhofer DEPARTMENT OF MEDICAL AND CLINICAL PSYCHOLOGY Dissertation Advisor	✓	
Dr. Cara Olsen DEPARTMENT OF PREVENTIVE MEDICINE AND BIOMETRICS Committee Member	✓	
Marjan G. Holloway, Ph.D. Dr. Marjan Holloway DEPARTMENT OF MEDICAL AND CLINICAL PSYCHOLOGY Committee Member	✓	

UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814

DISSERTATION APPROVAL FOR THE DOCTORAL DISSERTATION IN THE MEDICAL AND CLINICAL PSYCHOLGY GRADUATE PROGRAM

Title of Dissertation: "Evaluating a Novel Eye Tracking Tool to Detect Invalid Responding in Neurocognitive Assessment"

Name of Candidate: David Barry
Doctor of Philosophy Degree
May 7, 2014

DISSERTATION AND ABSTRACT APPROVED:

DATE:

Dr. Andrew Waters — 5/7/14
DEPARTMENT OF MEDICAL AND CLINICAL PSYCHOLOGY
Committee Chairperson

Dr. Mark Ettenhofer — 5/7/14
DEPARTMENT OF MEDICAL AND CLINICAL PSYCHOLOGY
Dissertation Advisor

Dr. Cara Olsen — 5/7/14
DEPARTMENT OF PREVENTIVE MEDICINE AND BIOMETRICS
Committee Member

Marjan G. Holloway, Ph.D. — 5-7-2014
Dr. Marjan Holloway
DEPARTMENT OF MEDICAL AND CLINICAL PSYCHOLOGY
Committee Member

ACKNOWLEDGMENTS

This dissertation project (and my graduate studies) would not have been possible without the tireless support of my wife Erin. Over the past four years, she has sacrificed her time and energy in order to give me every opportunity to succeed as a graduate student. She spent countless hours caring for our two children in order to give me time to complete this manuscript. She is an amazing teammate and I owe a great deal of my success to her unwavering support.

I was extremely fortunate to have joined Dr. Mark Ettenhofer's lab in 2010. We didn't have many resources at the time; the lab was just being set up and the eye-tracking equipment was still in boxes. In my first year, Dr. Ettenhofer spent a tremendous amount of his personal time and effort developing me as a researcher and clinician. He understood my strengths and targeted my deficiencies. He catalyzed my growth and challenged me to innovate. He trusted me. Together, we developed The Bethesda Eye & Attention Measure (BEAM), the eye-tracking program with exciting prospects for use in clinical and research settings. Coming from an engineering background, I enjoyed the development and refinement of a new system. Dr. Ettenhofer ensured the ship was steered towards the right directions. With his guidance, the BEAM and this project were made possible.

Doug Girard deserves ample credit for his role in programming the BEAM. I will be forever grateful that I met him at a graduate student networking event at a Baltimore Orioles game in 2010. Without his technical knowledge of E-Prime and coding structure,

Dr. Ettenhofer and I would not have been able to develop the BEAM software as quickly and efficiently as we did. Doug is a colleague and a friend.

I would like to thank the other members of my dissertation committee—Dr. Andrew Waters (chair), Dr. Marjan Holloway, Dr. Cara Olsen, and COL (Dr.) Rodney Coldren. Their guidance throughout the dissertation project cycle—project development, refinement, data collection, data analysis, interpretation, and presentation—was invaluable to my development as a scientist. They encouraged me and pushed me to achieve more than I thought I could with this project. Their responsiveness, collegiality, and wisdom made this dissertation project exciting and enjoyable.

Dr. Neil Grunberg deserves a special acknowledgement, as he has been an exceptional mentor for me throughout my time at USUHS. Before I even began my graduate studies, he showed me the standard he sets for his students and challenged me (expected me) to exceed it. In addition to the research skills and psychological knowledge he imparted upon me, Dr. Grunberg taught me a great deal about professionalism, teamwork, humility, and leadership. I will always appreciate everything he has done for my family and me.

I would like to thank the MPS faculty and staff who decided in 2010 to accept an engineer with minimal background in psychological sciences into their Clinical Psychology Ph.D. program. They took a chance on me, knowing that I was raw and needed considerable refinement. They and the outstanding adjunct professors (Dr. Claire DeLabar's Assessment I, II, III courses especially) gave me the training and knowledge I needed to immerse myself into a new field. I would especially like to thank the late Corinne Simmons for her friendship. She will be missed.

The project described in this manuscript was a true team effort, and I want to thank the graduate students and research assistants of the Ettenhofer Laboratory for Neurocognitive Research for their help recruiting participants, collecting data, and helping with administrative procedures. Special thanks to Dmitry Mirochnitchenko, our former lab manager, for all of his help with BEAM projects over the years. He was essential for technical support, troubleshooting, and administrative support. I would also like to extend a special “thank you” to Jessica Kegel, Vasiliki Anagnostopoulos, Miriam Littmann, Kathy Williams, and Amanda Devane for their efforts in getting this project completed.

I will be forever thankful to the support from my friends in the program: those in the cohorts above me, those in the cohorts below, and my amazing quals-prep-road-trip-rafting-adventure-DC-racing-PX-teammates. You all helped me more than you know. To the countless others who have helped me complete this project…thank you.

Copyright Statement

The author hereby certifies that the use of any copyrighted material in the dissertation manuscript entitled "Evaluating a Novel Eye-Tracking Tool to Detect Invalid Responding in Neurocognitive Assessment" is appropriately acknowledged and, beyond brief excerpts, is with the permission of the copyright owner.

David M. Barry

ABSTRACT

Evaluating a Novel Eye-Tracking Tool to Detect Invalid Responding in Neurocognitive Assessment

David M. Barry, Ph.D., 2015

Thesis directed by: Mark Ettenhofer, Ph.D., Assistant Professor, Department of Medical and Clinical Psychology

INTRODUCTION: Valid symptom report and test performance are essential prerequisites for the accurate interpretation of neurocognitive or neuropsychological assessment data. Unfortunately, base rates of invalid responding in civilian and military contexts suggest that symptom exaggeration and underperformance are common in these types of assessments. Many response validity tests (RVTs) have been developed and derived to detect invalid responding, but these measures are limited by lengthy administration times, limited sensitivity, and susceptibility to coaching.. This dissertation project evaluated a novel eye-tracking tool, the Bethesda Eye & Attention Measure (BEAM), as a method for detecting invalid responding in neurocognitive assessment.

METHODS: A prospective, simulator study compared neurocognitive battery performance between two group of healthy adults: an unbiased group (n=26) instructed to perform their best and a biased group (n=24) instructed to simulate deficits associated with head injury. The biased group was given a warning to fake believably. Results

from the simulator study were cross-validated in a clinical sample of unbiased responders with a history of mild TBI (n=19).

RESULTS: Of the 29 BEAM metrics evaluated in the simulator study, 12 demonstrated outstanding classification accuracy (AUC ≥ .90). Overall Saccadic Reaction Time Intra-Individual Variability (AUC = .97) and Overall Manual Reaction Time Intra-Individual Variability (AUC = .97) demonstrated the best classification accuracy among the BEAM variables. The BEAM performed favorably when compared to well-validated embedded and freestanding response validity tests—including the CPT-II, WAIS-IV Digit Span, Trail Making Test A & B, MSVT, and VSVT. Several BEAM metrics identified in the simulator study demonstrated outstanding classification accuracy in the clinical sample.

DISCUSSION: The BEAM demonstrated considerable promise as a tool to detect invalid responding in neurocognitive assessment. Consistent with the literature on continuous performance tests, BEAM reaction time intra-individual variability, omissions, and commissions demonstrated the best classification of invalid responding behavior in both experimental and clinical samples. This study adds to the extant response validity literature by demonstrating that saccadic performance in a continuous performance test may be used to detect invalid responding. Results from the simulator study were cross-validated in a clinically-relevant population, providing preliminary evidence supporting the BEAM's clinical utility as a response validity test.

Additional research should evaluate the BEAM's ability to identify invalid responding in larger, more heterogeneous groups of persons with and without neurological conditions.

TABLE OF CONTENTS

LIST OF TABLES xii

LIST OF FIGURES xiii

LIST OF SUPPLEMENTAL MATERIALS xiv

LIST OF ACRONYMS AND ABBREVIATIONS xv

CHAPTER 1: Introduction 1

- Introduction and Scope of Problem 1
  - Traumatic Brain Injury and Validity Concerns 5
  - Costs of Undetected Symptom Exaggeration and Invalid Responding 8
- Relevant Terminology Used in the Extant Literature 9
  - Describing Invalid Responding Behavior 9
    - Malingering: A Judgmental Description 11
    - Effort: An Insufficient Description 14
    - Negative Response Bias and Noncompliance: Newer Descriptions, Similar Concepts 16
  - Describing Validity Testing and Assessment 16
  - Proposed Terminology Changes 18
- Research Designs 21
- Base Rates of Invalid Responding 26
  - Base Rates of Invalid Responding in Civilian Populations 26
  - Base Rates of Invalid Responding in Active Duty U.S. Service Member and Veteran Populations 29
  - Base Rates Summary 35
- Diagnostic Validity and Classification Accuracy Statistics 36
- Existing Types of Symptom Validity Tests 40
  - Freestanding Symptom Validity Tests 41
  - Embedded Validity Indices 44
  - Limitations of Freestanding SVTs and Embedded Validity Indices 47
- Reaction Time and Embedded Validity Indices in Continuous Performance Tests 48
- Attention and Oculomotor Functioning 51
  - Fixations and Saccades 52
  - Saccadic Processes and Cognition 53
  - Eye Tracking Research of Cognitive Functioning 55
- The Bethesda Eye & Attention Measure (BEAM) 56
- Scope of Project 59
- Specific Aims and Hypotheses 60

CHAPTER 2: Methods 65

- Study Design 65
- Participants 66
  - BR and UR Groups 67
  - UR-mTBI Group 67
- Setting and Equipment 69

Setting 69
Computers 70
Eye-Tracking Device 70
Response Pad 71
Software 71
Data Acquisition and Post-Processing 71
Data Acquisition Procedure 71
Data Post-Processing 72
Independent Variables 73
Dependent Variables (Measures) 74
Sample Characterization Measures 75
Baseline Interview 75
Feedback Interview 75
Wechsler Test of Adult Reading (WTAR) 76
Head Injury Knowledge Scale (HIKS) 76
Oculomotor Measure 77
Bethesda Eye & Attention Measure (BEAM) 77
Cognitive Performance Tasks with Embedded Response Validity Indices 78
Conners' Continuous Performance Test, Second Edition (CPT-II) 78
WAIS-IV Digit Span Subtest 82
Trail Making Test (TMT) 85
Freestanding Response Validity Tests 86
Victoria Symptom Validity Test (VSVT) 86
Medical Symptom Validity Test (MSVT) 87
Procedure 88
Simulator Study: BR and UR Groups 89
Parent Study: UR-mTBI Group 91
Data Analysis 92
General Analytic Strategy 92
Control Variables 93
Data Analytic Strategy for Aims and Hypotheses 94
Specific Aim 1 94
Specific Aim 2 95
Specific Aim 3 97
Alpha Level 98
Power Analysis 99
CHAPTER 3: Results 100
Specific Aim 1 100
Specific Aim 2 106
Embedded RVTs 106
Freestanding RVTs 109
Hierarchical Logistic Regressions 112
Specific Aim 3 114
CHAPTER 4: Discussion 117
Summary and Integration of Results 117
Explanations for Findings 120

Limitations 132
Future Directions 136
Conclusion 137
APPENDIX A: TABLES 139
APPENDIX B: FIGURES 160
REFERENCES 200

LIST OF TABLES

Table 1: Contingency table for response validity test outcomes in a TBI population... 139
Table 2: Embedded response validity test cutoff scores used for UR-mTBI group 139
Table 3:Demographic characteristics of simulator study groups..... 140
Table 4: ROC analyses for BEAM variables in simulator study (UR vs. BR groups)... 141
Table 5: Non-normally distributed BEAM variables with outstanding classification accuracy 142
Table 6: Normally distributed BEAM variables with outstanding classification accuracy 142
Table 7: Cumulative percentages of persons with scores above the indicated cutoff on selected BEAM variables..... 143
Table 8: ROC analyses for embedded RVT variables in the simulator study 146
Table 9: Non-normally distributed embedded RVT variables with outstanding classification accuracy 146
Table 10: Normally distributed embedded RVT variables with outstanding classification accuracy 146
Table 11: Cumulative percentages of persons with scores above or below the indicated cutoff on selected embedded RVT variables 147
Table 12: ROC analyses for freestanding RVT variables..... 148
Table 13: Non-normally distributed freestanding RVT variables with outstanding classification accuracy 148
Table 14: Cumulative percentages of persons with scores above or below the indicated cutoff on selected freestanding RVT variables..... 149
Table 15: Hierarchical logistic regressions examining incremental validity of representative variables..... 151
Table 16: Demographic characteristics of clinical comparison group..... 152
Table 17: Group comparisons of normally distributed BEAM and embedded RVT variables with excellent classification accuracy 153
Table 18: Group comparisons of non-normally distributed BEAM and embedded RVT variables with excellent classification accuracy 154
Table 19: Cumulative percentages of persons with scores above the indicated cutoff on selected BEAM variables..... 155
Table 20: Cumulative percentages of persons with scores above or below the indicated cutoff on selected embedded RVT variables 159

LIST OF FIGURES

Figure 1: Stepwise Logistic Regression Illustration for BEAM Variables..... 160
Figure 2: Variable Reduction Depiction for Simulator Study Variables..... 160
Figure 3: Saccadic RT-IIV-Directional Cue Sensitivity and Specificity by Cutoff Score 161
Figure 4: Manual Omission Error Rate Sensitivity and Specificity by Cutoff Score..... 161
Figure 5: Manual RT-IIV-Misdirectional Cue Sensitivity and Specificity by Cutoff Score 162

LIST OF SUPPLEMENTAL MATERIALS

Picture 1: Computer Monitor, ASL D6 Eye Tracker, and Cedrus Response Pad 163
Picture 2: ASL EYE-TRAC 6 Control Unit and Computers 163
Picture 3: Examiner Station with ASL LCD Monitors 164
Script 1: Simulator Study Group Assignment Scripts 165
Script 2: Simulator Study Debrief Scripts 167
Measure 1: Simulator Study Feedback Interview 168
Measure 2: Simulator Study Baseline Interview 170
Measure 3: Parent Study TBI Cohort Baseline Interview 175
Administrative Document 1: Simulator Study Informed Consent Form 188
Administrative Document 2: Parent Study TBI Cohort Informed Consent Form 193
Administrative Document 3: Simulator Study Recruitment Flyer 198
Administrative Document 4: Parent Study TBI Cohort Recruitment Flyer 199

LIST OF ACRONYMS AND ABBREVIATIONS

AACN: American Academy of Clinical Neuropsychology

ACSS: Age-corrected scaled score

APA: American Psychological Association

ASL: Applied Science Laboratories

ARE: Asymptotic relative efficiency

AUC: Area under the curve

BEAM: Bethesda Eye & Attention Measure

BR: Biased responding group (invalid responding/simulated TBI group)

CPT-II: Conners' Continuous Performance Test-II

DS: Digit Span

DSM: Diagnostic and Statistical Manual of Mental Disorders

HIKS: Head Injury Knowledge Scale

HR: Hit rate

K-D Test: King-Devick Test

IRB: Institutional Review Board

ManRT: Manual Reaction Time

ManRT-IIV: Manual Reaction Time Intra-Individual Variability

ManCom%: Manual Commission Error Percentage

ManOm%: Manual Omission Error Percentage

MND: Malingered Neurocognitive Dysfunction/Malingered Neuropsychological Dysfunction

MPRD: Malingered Pain-Related Disability

MSVT: Medical Symptom Validity Test

NAN: National Academy of Neuropsychology

NSI: Neurobehavioral Symptom Inventory

NPV/NPP: Negative predictive value/negative predictive power

PPV/PPP: Positive predictive value/positive predictive power

PVA: Performance validity assessment

PVT: Performance validity test

RDS: Reliable Digit Span

ROC: Receiver operating characteristic

RVA: Response validity assessment

RVT: Response validity test

SacRT: Saccadic Reaction Time

SacRT-IIV: Saccadic Reaction Time Intra-Individual Variability

SacCom%: Saccadic Commission Error Percentage

SacOm%: Saccadic Omission Error Percentage

SN: Sensitivity

SP: Specificity

SPV: Symptom/performance validity

SPVA: Symptom/performance validity assessment

SPVT: Symptom/performance validity test

SSA: Social Security Administration

SVA: Symptom validity assessment

SVT: Symptom validity test

TBI: Traumatic brain injury

TMT: Trail Making Test

UR: Unbiased responding group (valid responding group)

UR-mTBI: Unbiased responding group with a history of mild traumatic brain injury

UCMJ: Uniformed Code of Military Justice

USUHS: Uniformed Services University of the Health Sciences

VA: Department of Veterans' Affairs

VDC: Veterans' Disability Compensation program

VSVT: Victoria Symptom Validity Test

WTAR: Wechsler Test of Adult Reading

CHAPTER 1: Introduction

INTRODUCTION AND SCOPE OF PROBLEM

Invalid responding, symptom exaggeration, malingering, faking behaviors, and other forms of inaccurate representations of abilities and conditions have transcended scientific and clinical disciplines for millennia. In their excellent piece on the history of deliberately misrepresented illnesses, Carone and Bush (44) provide several examples of this phenomenon throughout recorded history: David from the Hebrew Bible acting erratically and drooling on his beard to escape persecution, Odysseus pretending to be insane to avoid fighting in the Trojan War, and notes from a 2nd century Greco-Roman physician documenting the simulation of pain and injury to avoid responsibilities. From the ancient ages to the present, humans have presented with invalid physical and psychiatric problems, often times doing so to achieve a desired result. The implications of such behaviors are not trivial; misrepresented physical and psychiatric conditions can influence medical treatments, return-to-duty evaluations, social expectations, and other clinical outcomes across a number of settings.

Since the early 1990s, the scientific field of neuropsychology has taken an increasingly active role towards understanding feigned, exaggerated, or otherwise invalid symptom and ability presentation. Neuropsychology is a clinical and experimental branch of psychology that studies the structure and function of the brain in relation to behaviors, emotions, cognitions, physical capacities, and symptom presentations (116). In turn, neuropsychologists aim to study, assess, and treat behaviors directly related to brain functioning by administering comprehensive evaluations on human subjects of interest. Once the assessments are scored, neuropsychologists can form diagnostic

impressions and clinical inferences regarding a person's physical, behavioral, emotional, and cognitive functioning and prognosis. In essence, neuropsychological assessment is a quantitative, standardized means of measuring complex aspects of human behavior and cognition, such as attention, memory, visuospatial and perceptual skills, language, reasoning, planning, and emotional processing (163).

A central principle of neuropsychological or "neurocognitive" assessment predicates a relationship between performance on neuropsychological tests and the actual condition of the brain (164). As such, neuropsychological assessment depends on the examinee's full effort and accurate symptom report for the results to be *valid* (154). If the results of a given neuropsychological assessment tool are invalid, the results cannot by definition be related to brain function. Knowing whether test data is valid is essential for drawing conclusions, making diagnoses, and recommending treatments (163). As such, it is imperative for examiners to consider both the psychometric properties of the measures used in neuropsychological assessment (e.g., internal factors; 87) as well as factors which may influence performance, such as environmental effects (181) or rapport with examiner (104).

This manuscript focuses on a particular factor that influences test outcome—response validity, or the validity of one's performance and symptom presentation during neurocognitive assessment. Neurocognitive assessment depends on its examinees to provide accurate symptom report and adequate level of effort to perform well throughout the testing process (132; 154). Knowingly or unknowingly, however, some persons undergoing neurocognitive assessment may give misleading responses or perform at levels other than their actual neurocognitive status (163). An individual's response

validity can be impacted by multiple patient factors, such as pursuit of secondary gain, fatigue, stress, medical conditions, psychiatric conditions[1], and medications, as well as external factors, such as testing environment (e.g., limited space, excessive ambient noise), examiner skill, unclear assessment instructions, and language/cultural barriers (11; 181).

Invalid presentations on neurocognitive assessment are not entirely attributable to neurological conditions, are not significantly influenced by demographic variables or performance confounds (e.g., fatigue, medication), and are significantly worse than expected scores for persons with genuine brain disorders (116). Beyond invalidating test results, invalid responding behavior can lead to significant individual, economic, and societal consequences, such as undetected neurological problems, improperly awarded financial settlements, and overly compensated disability claims (215). Unfortunately, symptom exaggeration and insufficient effort to perform well are frequently identified in neuropsychological evaluations, especially in forensic settings (34; 41; 154).

Invalid responding behavior can manifest on performance tests of abilities (e.g., cognitive abilities, motor abilities) and self-report measures (e.g., symptom presentation measures). Unfortunately, subjective methods for detecting response bias, such as clinical judgment (i.e., clinical intuition), pattern analysis, and discrepancy methods, are insufficient and influenced by cognitive biases (106; 115). To more accurately detect invalid performance and exaggerated symptom presentation, clinicians utilize a variety of stand-alone validity tests and validity indices embedded within assessment measures. When persons surpass validity thresholds on these measures or indices, clinicians have objective evidence to suggest the presence of invalid data.

[1] Somatoform or cogniform disorders

Largely in part to a surge of interest and research in response validity in the 1990s (254), researchers were able to estimate base rates of deliberate misrepresentations of abilities or symptoms in a host of civil and criminal settings (154). Several decades of research suggest that symptom exaggeration and response bias occur on 30-50% of neuropsychological evaluations with potential for secondary gain (158; 183). This high rate of suboptimal performance is even more disturbing when considering that one's effort to perform optimally on neurocognitive tests accounts for more variance than brain injury severity (82; 95; 99; 150; 178). In the past decade, the National Academy of Neuropsychology (NAN) and the American Academy of Clinical Neuropsychology (AACN) recommended response validity assessment be included in all evaluations of brain and behavior (41; 116). The NAN paper added that effort and symptom validity assessment could be conducted with specially-designed tests, indices, and observations of effort, as well as other non-specific metrics (41).

Researchers have been evaluating methods of response bias detection for decades in order to adapt to the ever-evolving patterns and presentations of response bias (218). A common research design used to study invalid responding is the "simulator study," where groups of participants instructed to perform their best are compared to groups of participants instructed to simulate deficits (248; 272). Simulation studies, also known as "analog research on dissimulation," allow for an experimental design where subjects can be randomly assigned to dissimulating or nondissimulating conditions (154). Since neurocognitive assessment often deals with neurological injury or illness, simulation study designs frequently utilize actual and simulated traumatic brain injury (TBI) groups for comparisons to healthy controls (23; 249-251; 279; 282). TBI assessment,

particularly with mild TBI, often relies on self-report and is thus vulnerable to response bias (242; 266; 286). This diagnostic limitation makes TBI groups an attractive population for studying feigned cognitive deficits.

Computerized, continuous performance tests have been used to study attention, executive functioning, and information processing in TBI populations (33; 54). Continuous performance tests have recently been evaluated as performance validity measures, with metrics such as button press reaction time, button press reaction time variability, omission errors, and commission errors showing acceptable sensitivity towards invalid responding (42; 118; 119; 149; 198). Oculomotor metrics and eye tracking methodologies have also been identified as promising tools for detecting invalid responding on neurocognitive measures (111).

This dissertation project investigated a novel invalid responding detection method that combines continuous performance tests with oculomotor functioning. To provide sufficient context for this study, the following literature review will present the problems associated with invalid symptom and ability presentation, describe the terminology used throughout the literature, and address the strengths and limitations of the current methods used to detect invalid responding. Next, the literature review will describe the potential for reaction time and oculomotor assessment to detect invalid responding. Lastly, a novel, multimodal neurocognitive assessment tool—the Bethesda Eye & Attention Measure (BEAM) will be described.

Traumatic Brain Injury and Validity Concerns

Traumatic brain injury (TBI) presents a significant public health and economic concern in the United States and the world. In 2009, there were approximately 3.5

million TBIs recorded in U.S. medical settings (56). While the vast majority (~70-90%) of treated TBIs in the U.S. are classified as "mild" (i.e., concussion; 22; 45; 47), the overall impact of TBI is anything but mild. According to population-based data obtained between 2002-2006, TBI contributes to nearly one-third of all injury-related deaths in the United States (76). TBI is associated with cognitive, psychological, physical, and behavioral sequelae which often lead to disability (131; 237). The total lifetime costs of all the fatal, hospitalized, and nonhospitalized cases of TBI that were medically treated in 2000 were estimated to be $60.4 billion (including productivity losses of $51.2 billion), with per-person lifetime costs approaching $45,000 (57). More recently, McCrea (171) estimated TBI to have a $100 billion annual impact on the U.S. economy in terms of medical costs and lost productivity.

Military-related TBI has dramatically added to the U.S. economic cost in the past decade, primarily due to the wars in Iraq and Afghanistan. It has been estimated that approximately 15-20% of all U.S. Service Members who deploy to Iraq or Afghanistan sustain at least one mild TBI (123; 258; 260). Using a standard cost-of-illness approach, the RAND Corporation (258) estimates the average cost in 2005 dollars of a deployment-related TBI to the U.S. economy ranges from $148,573 to $222,000 per TBI. The total U.S. economic cost of deployment-related TBI from 2001-2005 is estimated to be between $90,629,389 to $135,419,773 (258). The median annual cost for TBI-diagnosed OIF/OEF Veterans was nearly four times higher than OIF/OEF veterans without a history of TBI (259). From January 1, 2000 through the fourth quarter of 2013, there have been 294,172 documented cases of all-severity TBI among Department of Defense Service Members (62). Of those, 242,676 (82.5%) were classified as mild TBI, 23,754 (8.1%) as

moderate TBI, 4,389 (1.5%) as penetrating, 2,920 (1.0%) as severe TBI, and 20,433 (6.9%) were not classifiable (62).

A recent meta-analysis of epidemiological studies of TBI in developed countries estimates lifetime prevalence of at least one TBI with loss of consciousness (LOC) in the general population to be 12%, with 16.7% lifetime prevalence for males and 8.5% for females (86). Each year in the U.S., more than a million people receive emergency room treatment for TBI, with 235,000 eventually being hospitalized and 50,000 dying from their head injuries(152). It is likely that the true incidence and prevalence of TBI-related disability is higher than these estimates, as the numbers do not incorporate unreported TBI or TBI treated outside of civilian hospitals (57). Recently, a population-based incidence study of TBI in New Zealand incorporated registered *and* nonregistered cases of TBI in their estimates, and the authors reported mild TBI incidence of 749 (95% *CI*: 709-790) per 100,000 person-years, a much higher incidence estimate than previously reported in studies from other modern countries (77).

With advances in modern medicine and neuroimaging, more civilians and service members are surviving TBI. As a result of reduced mortality rates, an ever-increasing number of people are living with major functional and cognitive disabilities (171). Between 3.17 and 5.3 million U.S. citizens (roughly 10% of all disabled Americans) are estimated to be living with permanent TBI-related disability (152; 262; 285). An estimated 43.3% of Americans have residual disability one year following TBI-related hospitalization (237).

Traumatic brain injury, like other disabling injuries or illnesses, has a profound impact on the U.S. economy. Unfortunately, TBI and other conditions that often rely

primarily on patient self-report for diagnosis are especially vulnerable to patient misrepresentation, provider misclassification, and improper disability compensation (242; 266; 286). While this diagnostic dilemma may prove difficult in many clinical and research settings, it makes TBI—and mild TBI in particular—an ideal population for studying feigned deficits (100).

Costs of Undetected Symptom Exaggeration and Invalid Responding

Symptom exaggeration and feigned disability can significantly impact the U.S. economy if undetected. It is estimated that 30-50% of all disability compensation evaluations involve some form of symptom exaggeration and/or invalid responding (158; 183), and that the total annual cost of insurance fraud to the U.S. economy approximates $85.3 billion (166). In 2008 alone, the Social Security Administration (SSA) and other governmental programs spent approximately $428.5 billion in payments to working-age persons who met disability criteria (a sizable increase from the $280 billion spent in 2002; 51). Of that $428.5 billion, Chafetz (51) estimates that $42.85 to $180 billion (i.e., 10-42% of total expenditures) were spent on claimants with possible, probable, or definite misrepresentations of disability.

Invalid representation of abilities or symptoms also poses a significant problem for the Department of Veterans' Affairs (VA), the agency that oversees the Veterans' Disability Compensation (VDC) program. In fiscal year 2004, nearly 2.5 million Veterans (10.2% of the total U.S. Veteran population) received disability compensation, with the average annual disability compensation payment being $8,378 (143). Applying Chafetz's (49; 51) base rates of symptom exaggeration and underperformance, between

$2.1 billion and $8.8 billion was likely spent in fiscal year 2004 on VA disability compensations involving exaggerated symptoms or underrepresented abilities.

It is clear that symptom exaggeration and ability misrepresentation in disability compensation contexts constitute major problems on a national economic level. Federal and state programs designed to support disabled civilians and Veterans may be improperly compensating individuals by tens of billions of dollars each year. As budgets at all levels of government are being scrutinized for waste, fraud, and abuse, it is clear that greater emphasis is needed towards detecting invalid responding in disability contexts. Undetected symptom exaggeration and invalid responding creates a significant financial and societal burden. Alternatively, implementing effective, evidence-based methods to detect invalid responding in federal disability programs could reduce this burden and save millions, perhaps billions, of dollars each year. Given the potential cost-savings, additional research on invalid responding behavior and detection is needed.

RELEVANT TERMINOLOGY USED IN THE EXTANT LITERATURE

Describing Invalid Responding Behavior

Proper research on response validity and its assessment first requires a method to operationalize constructs (28). Unfortunately, a plethora of loosely defined words, terms, and definitions have been used to describe similar constructs throughout the literature on response validity. One of the most common (and controversial) terms used in the literature is *malingering*, which was derived from the French word *malingre*, meaning "sickly" (174). Another popular term, *effort*, has been utilized to indicate the amount of mental and/or physical energy expended in performing a task at capacity levels (41; 99; 246). *Negative response bias* and *dissimulation* have been also been used to describe the

misrepresentation of abilities, either from over-representing or under-representing a true set of symptoms during evaluation (41; 116).

Not surprisingly, several terminology "camps" have emerged. While some researchers staunchly defend use of the term *malingering* or some derivation thereof (244), others advocate using alternate terms such as *suspect effort* (14) or *incomplete effort* (12). Boone (34) prefers terminology that describes invalid responding behavior irrespective of intent, such as *noncredible performance/symptoms, negative response bias,* or *non-physiological, suspect, or suboptimal effort*. This nomenclature dilemma stems from an important debate involving the ever-evolving construct of malingering, the intentionality of suboptimal performance and presentation, and what the definition of "effort" actually means from a biopsychosocial perspective (29; 30; 156; 246).

As a result of the terminology debates, the list of terms associated with response validity in the context of symptom presentation, effort, and performance has grown into a veritable thesaurus. In addition to the aforementioned terms, it is not uncommon to find *feigned cognitive impairment, nonorganic signs and symptoms, insufficient effort, invalid effort, invalid/failed performance, cognitive malingering, faking bad, symptom amplification, performance exaggeration, underperformance/distortion, symptom embellishment, disingenuous,* or *faked* in a neurocognitive evaluation report (28; 30; 34). Bigler (30) argues that scientists have created a tautological problem of unnecessary and repetitive use of different words, terms, and acronyms with similar meanings, creating communication barriers for researchers and clinicians across scientific disciplines.

To adequately review the literature of symptom exaggeration and response validity in neurocognitive test performance, it is necessary to first provide a review of the

terminology used to describe it. The following section will present relevant constructs and their definitions, including *malingering*, *effort*, and other terms used to describe validity assessment measures. The intent of this section is to clarify the similarities and differences among common terms in the response validity literature, and to arrive upon a consistent terminology and construct operationalization for use in this manuscript.

Malingering: A Judgmental Description

Merriam-Webster (174) defines malingering as the pretending or exaggeration of incapacity or illness as to avoid duty or work. In the U.S. military, malingering is a punishable offense under Article 115 of the Uniformed Code of Military Justice (UCMJ; 206). According to UCMJ (211, np), malingering is done for the "purpose of avoiding work, duty, or service," and manifests as the (1) intentional infliction of self-injury, or (2) feigning of "illness, physical disablement, mental lapse, or derangement."

The American Psychiatric Association (APA) listed the term *malingering* in the original *Diagnostic and Statistical Manual of Mental Disorders* (DSM; 2) as a supplemental term in an appendix without specific criteria. In *DSM-II* (3), malingering was described as a conscious behavior that needed to be distinguished from Hysterical Neurosis, Conversion Type. In 1980, *DSM-III* listed malingering as a "V" code—a condition not considered to be a mental disorder *per se*, but still worthy of clinical attention. The original *DSM-III* malingering criteria were the presence of false or exaggerated physical or psychological symptoms, voluntarily produced in the pursuit of an obvious, recognizable goal. Minor changes were made to the malingering criteria in *DSM-III-R* (5), *DSM-IV* (6) and *DSM-IV-TR* (7), and the criteria appear to remain the same for *DSM-5* (26). *DSM-IV-TR* (7) indicates that malingering may represent an

adaptive behavior and recommends strong consideration of malingering if one or more of the following is present: (1) medicolegal context, (2) marked discrepancy between objective findings and a person's symptom report, (3) poor treatment compliance or rapport with provider, and (4) the presence of antisocial personality disorder. Of note, the *DSM-IV-TR* (7) states that malingering must be distinguished from both factitious disorder (i.e., conscious symptom generation to fulfill the "sick role") and somatoform disorder (i.e., unconscious symptom generation).

While the *DSM*'s definition of malingering has changed somewhat over the past several editions, ample evidence suggests its diagnostic criteria are clinically and practically untenable for use in clinical practice and research (26). To provide a more reliable framework for malingering in clinical and research settings, particularly in the field of forensic neuropsychology, Slick, Sherman, and Iverson (244) introduced differential diagnostic criteria for Possible, Probable, and Definite Malingered Neurocognitive Dysfunction (MND). This set of criteria has since become the most commonly used diagnostic standard in neuropsychological research (246). MND is defined as "the volitional exaggeration or fabrication of cognitive dysfunction for the purpose of obtaining substantial material gain" (e.g., financial compensation for injury), "or avoiding or escaping legally obligated formal duty" (e.g., child support payments, military deployments) "or responsibility" (e.g., competency to stand trial; 244; p. 552).

Borrowing heavily from the 1999 Slick et al. criteria, Bianchini, Greve, and Glynn (28; pg. 407) introduced Malingered Pain-Related Disability (MPRD) as the "intentional exaggeration or fabrication of cognitive, emotional, behavioral, or physical dysfunction attributed to pain for the purposes of obtaining financial gain, to avoid work,

or to obtain drugs (incentive)," and provided differential diagnostic criteria for Possible, Probable, and Definite MPRD. The authors proposed that compelling inconsistencies (i.e., unambiguous discrepancies) between examinee symptom report, test performance, and/or behavior should be considered pathognomonic of malingering when external incentives are present and the behaviors are not better accounted for by legitimate neurological or psychiatric disorders. Unlike the 1999 Slick et al. criteria, MPRD incorporates evidence from a physical evaluation into its criteria.

After more than a decade since first operationalizing malingered neurocognitive dysfunction, Slick and Sherman (245) recently published a set of revised diagnostic criteria for what they now call Malingered *Neuropsychological* Dysfunction (same MND acronym; italics added for emphasis). Instead of Possible, Probable, or Definite subtypes of one diagnosis, MND, Slick and Sherman (245) propose Probable or Definite subtypes of three *separate* diagnoses: Primary MND, Secondary MDN, and MND by Proxy. Primary MND is diagnosed when external incentive is present, when exaggeration/fabrication of neuropsychological problems or deficits is detected, and when behaviors are not substantially accounted for by a psychiatric, neurological, or developments factors (245; 246). Secondary MND incorporates the possibility of diminished cognitive capacity and/or inability to control one's behavior due to legitimate, severe cognitive/psychiatric dysfunction (e.g., severe TBI, schizophrenia, or mental retardation; 245; 246). MND by Proxy is diagnosed when minors meet criteria for malingering primarily as a result of the intentional influence or control of an adult (245; 246).

While multiple empirically-supported diagnostic criteria for malingering exist, researchers and clinicians often avoid using the term *malingering* due to its pejoratively judgmental connotation and potential legal ramifications, preferring "softer" terms instead (132; 227). In fact, the Social Security Administration (SSA) admonishes its psychologists not to use the term *malingering* because it is "too subjective" (50). *Malingering* and its related terminology remain highly controversial, mainly because malingering, by definition, is an intentional process (246).

Proponents of malingering diagnoses believe this deceitful intent can be inferred through rigorous empiricism, calling the practice of substituting "softer" terms to describe malingering unethical per American Psychological Association (APA) guidelines that require findings to be presented as unambiguously as possible (244). Opponents of malingering diagnoses believe intent *cannot* be reliably inferred from cognitive tests, and malingering (a conscious behavior) cannot be reliably differentially diagnosed from conversion, factitious, or major depressive disorders (conscious and unconscious behaviors; 35; 132). Boone (35) argues that the term *malingering* should only be used in the rare circumstance where there is incontrovertible evidence of malingering (e.g., patient admission, surveillance footage), and recommends changing the 1999 Slick et al. terminology from "diagnosis of malingered neurocognitive dysfunction" to "determination of noncredible neurocognitive function."

Effort: An Insufficient Description

Neuropsychological assessment depends on the examinee's full effort and accurate symptom report for the results to be valid (154). The term *effort* describes the amount of mental and/or physical energy expended in performing a task (246), and is

often used to describe a subject's investment in performing at his or her capacity levels (41). In the context of neuropsychological evaluations, the term *effort* implies a unidirectional drive to respond in a valid manner. While malingering is characterized by the goal to which effort is directed (i.e., secondary gains), effort is often characterized by *level* of investment towards performing well. For example, it is common to see the term *poor effort* used to describe persons who do not attempt to do their best (99). Unlike the term *malingering*, effort does not imply intent or motivation, but rather serves to describe responding behavior in general.

Unfortunately, the term *effort* is often loosely and inappropriately applied in clinical and research settings, which confounds its meaning within and across disciplines (246). For example, the qualitative and quantitative components of effort have different meanings in biology, cognitive neuroscience, and neuropsychology (30; 228). From a neuropsychologist's perspective, the term *effort* does not differentiate between people who are not trying hard enough from those who are deliberately attempting to deceive. While failure on an *effort test* could reflect inadequate or poor effort to perform well, it could also signify considerable effort towards performing poorly! As such, *poor effort*, *inadequate effort*, *suboptimal effort*, and other aforementioned effort descriptors may inappropriately characterize invalid responding behaviors. Slick and Sherman (246) recommend clinicians use more specific terms to avoid misunderstandings, such as *noncompliance* (see below). Bigler (30) argues that the term *effort* should be abandoned altogether.

Negative Response Bias and Noncompliance: Newer Descriptions, Similar Concepts

Other terms less common than *malingering* or *effort* are beginning to gain support in the response validity literature. In neuropsychology, *negative response bias* is characterized by an attempt to mislead examiners through inaccurate or incomplete responses (41) or the misrepresentation of abilities through performance or self-report (116). The "negative" component of the term refers to the negative valence of response bias that would be expected (i.e., poorer performance or worse symptom presentation). Like the terms *malingering* and *poor effort*, negative response bias is detected when persons fail to surpass thresholds of valid performance on validity measures or validity indicators within ability tests and/or self-report measures (116). Negative response bias describes the behavior without inferring intent (35).

Slick and Sherman (246) propose describing failed validity indicators as a form of *noncompliance* with test instructions, with *compliance* being defined as the attempt to complete tasks in accordance with the specific directions given for each test. Put another way, compliant patients attempt to respond correctly (84). Since the instructions for any given neurocognitive assessment explicitly state that persons should make every attempt to perform their best, examinees who fail to try their best would exhibit noncompliance. Noncompliance could occur consciously or unconsciously, and does not imply intent. Noncompliance with test instructions describes both poor effort to do well and maximal effort to perform poorly (246).

Describing Validity Testing and Assessment

Like the terms to describe the invalid responding behavior, there is no consensus terminology used to describe validity assessment. *Symptom validity assessment (SVA),*

symptom validity tests (SVTs), *effort testing*, *effort tests*, and *malingering tests* are common terms seen throughout the literature, and they are often used interchangeably. There are generally two types of validity tests or indicators: *freestanding* (i.e., *stand-alone*) measures that are administered independently of other measures, and empirically derived *embedded* indices, scores, or markers found within self-report or ability tests. The origin and evolution of these assessment terms deserve special consideration in order to understand their use in clinical and research settings.

Pankratz and colleagues (199; 200) originally coined the term *symptom validity test (SVT)* to describe a dichotemous, forced-choice paradigm that used below chance responding to detect exaggeration or faking of one's symptom presentation. Originally, the term was used to identify conversion disorder and rule out malingering (199). Many of the original freestanding validity assessment tools used the forced-choice paradigm, and the term *SVT* quickly became associated with forced-choice responding on memory tests of digit and letter/word recognition (163). Over time, however, the term *SVT* has become a sort of "catch-all" term that can be applied to any tool or index designed or derived to evaluate the validity of symptoms and test performance (106). The term *SVT* can describe freestanding measures of all formats (not just forced-choice paradigms) as well as embedded validity indicators within tests of cognitive ability and self-reported symptom inventories (51; 105; 178). Similarly, the terms *symptom validity assessment* and *symptom validity testing* have evolved to encompass assessment of both symptom exaggeration as well as invalid responding on ability tests.

Effort testing and *effort tests* are other common terms used alongside *symptom validity assessment* and *symptom validity tests* (99; 151; 175). Green (95; 99) routinely

refers to his freestanding measures as *effort tests*. Clinicians and researchers may prefer the term *effort tests* to describe methods to evaluate whether cognitive or physical effort is sufficient to produce valid data (96). Additionally, embedded validity indicators are sometimes called *embedded effort measures* (42).

As described earlier, the term *effort* implies a unidirectional motivation to perform at one's best. *Effort tests* and *effort testing*, accordingly, are purported to detect the incidence of poor or suboptimal effort to perform at capacity levels. By this logic, effort test failure should equate to *insufficient* effort. However, as Slick and Sherman (246) note, effort test failure could also indicate high levels of effort towards performing below one's actual capacity. As such, the term *effort testing* may be an insufficient description of what is being measured.

Proposed Terminology Changes

Clearly, the plethora of words and terms used to describe invalid responding behavior and the tools used to detect it presents communication challenges between clinicians, researchers, and the general public. Some terms, like *malingering*, are pejorative. The term *malingering* carries legal ramifications, and many clinicians are hesitant to label someone as a "malingerer." Other terms, like *effort*, are controversial. As it is currently used in the neuropsychology literature, the term *effort* insufficiently describes its own construct. Terms like *symptom validity assessment* and *symptom validity tests* are modern misnomers that have evolved beyond their original meanings (106).

Several proposals have recently been made to clarify the terminology. Slick and Sherman (246) suggest using *noncompliance detection measures* in lieu of *effort tests*

when describing response validity tools[2]. Bianchini, Curtis, and Greve (27) use the term *cognitive performance validity tests (CPVTs)* to describe freestanding measures or indices used to determine if an individual has underperformed on tests of perceptual and/or cognitive ability. Recently, Larrabee (156) and Bigler (30) each addressed invalid responding terminology issues and by consensus advocated abandoning the use of the term *effort* in favor of two distinct terms: *symptom validity* and *performance validity*. Under this framework, *symptom validity* would solely describe the accuracy of one's symptom presentation on self-report measures such as the Personality Assessment Inventory (PAI; 186), and *performance validity* would describe the veracity of ability task performance on freestanding measures or embedded validity indices derived from existing neurocognitive tests (156). Accordingly, *symptom validity tests (SVTs)* would be used to detect symptom exaggeration, and *performance validity tests (PVTs)* would detect underrepresentation of one's abilities.

The litany of terms used to describe the measures designed to detect invalid responding, the assessment of invalid responding, and invalid responding behavior presents a tautological problem for clinicians, researchers, and lay individuals (30). Several scholars have proposed interesting and conceptually convincing arguments towards a new vernacular, arguing for more specific terms like *performance validity* to describe ability representation and *symptom validity* to describe symptom presentation (30; 156). This author believes that the more specific these terms become, the less generalizable they are to settings outside of a given field. Instead of carving out niche words for specific uses (e.g., *cognitive performance validity tests*), this author argues that

[2] Presumably, "noncompliance detection assessment" would describe the assessment of noncompliance.

researchers and clinicians should opt for simpler, more global terms that accurately describe constructs. Despite their previously mentioned limitations, *symptom validity test* and *symptom validity assessment* have been adopted as "catch-all" terms largely due to their ease of use and universal application (106). In the field of neuropsychology, an *SVT* generally describes anything that is designed to detect invalid responding, regardless of whether symptom presentation is considered.

If the ultimate goal of modern symptom validity tests and symptom validity assessment is to determine *response validity*, then the terms *response validity tests (RVTs)* and *response validity assessment (RVA)* appear to be more appropriate. *Response validity tests* can describe self-report measures, performance measures, freestanding measures, or embedded indices without implying symptom involvement. *Response validity assessment* can describe the assessment of all types of invalid responding. The terms *response validity tests* and *response validity assessment* are simple, global, and, most importantly, accurate. These terms also avoid the eventual use of more cumbersome conceptual descriptions such as *symptom and performance validity (SPV), symptom and performance validity assessment (SPVA),* and *symptom and performance validity tests (SPVTs).*

The term *response validity assessment* also lends itself to impartial and nonjudgmental descriptions of invalid responding behavior. *Malingering, poor effort, noncredible responding*, *negative response bias*, and *noncompliance* all impart some sense of blame onto the examinee, deserved or not. *Invalid responding*, used frequently thus far in this literature review, can easily fit into the *response validity test/assessment* framework. For example, one or more failed response validity tests during a

neurocognitive assessment could raise suspicion of invalid responding, and the overall assessment's validity could be questioned. Consistent with Boone's (34) recommendations, *invalid responding* describes a behavior without implying intent.

While this author supports the use of the terms *response validity tests*, *response validity assessment*, and *invalid responding*, it would be imprudent in a literature review to retroactively apply these terms to describe previous research. Accordingly, the terms used throughout this dissertation when describing prior research will reflect the terms used in their respective sources, and care will be made to accurately convey the authors' intended meaning. For sections written specifically about this dissertation project, however, the terms *response validity tests*, *response validity assessment*, and *invalid responding* will be used.

RESEARCH DESIGNS

Given the drastic economic cost of undetected invalid responding in disability and legal compensation claims, researchers have sought to identify the prevalence and presence of invalid responding. In order to determine base rates of invalid responding and understand how invalid responding behavior manifests in various settings and clinical contexts, researchers have commonly used the following research designs: case studies, differential prevalence designs, simulation studies (i.e., analogue research on dissimulation), and known-group designs (155; 192; 217; 218). Each design carries unique methodological strengths and weaknesses. Case studies, for example, are useful for generating qualitative information and hypotheses for future research. Early malingering research using case studies (121; 200) eventually led to the development of the Portland Digit Recognition Test (PDRT; 31; 155). While the generalizability of case

study results is clearly limited by a small sample size and other methodological weaknesses, a case study design is often the only practical design for research of rare syndromes, diseases, or conditions (155).

Differential prevalence designs allow researchers to describe outcome frequencies in a given population without systematically investigating the cause of the differences (192). Researchers using a differential prevalence design infer *a priori* that two or more samples will have a different prevalence of a condition (e.g., clinical referrals for lung problems compared to injury litigants). Differential prevalence designs have been used to describe the effect sizes of varying levels of financial compensation and neuropsychological assessment performance (32).

Differential prevalence designs, like case studies, are significantly limited in their internal and external validity. The designs by their nature do not incorporate independent criteria for the topic of interest (i.e., feigned performance); as a result, researchers cannot determine who or how many in each group are dissimulating. The design allows analysis for only overall differences based on assumptions that groups will have different rates of malingering (155; 192; 218). While Rogers (218) argues that the differential prevalence design is the weakest methodology for symptom validity assessment research, others contend the design can be useful for grasping a preliminary understanding of how symptoms present in understudied clinical populations (192). The differential prevalence design, like correlational studies, are useful for identifying group phenomena that merit further exploration using more rigorous scientific methods.

Simulation or “analogue malingerer” research employs a quasi-experimental design where subjects can be assigned randomly to different scenario groups (218).

Typically, one group is instructed to exaggerate symptoms or feign impairment related to a condition of interest. This "biased" or "dissimulator" group is compared to another group of randomly-assigned subjects who are instructed to perform normally or "honestly" (192). The biased groups are often given a real or pretend incentive (e.g., college credit, monetary compensation, etc.) to enhance motivation to perform like a "real-world malingerer." Researchers can use this design with clinical groups to determine effects of invalid responding above and beyond the effects of the clinical condition (e.g., ADHD, mild TBI, etc.).

Simulation study designs allow researchers to better understand how invalid symptom expression and performance manifests on a given assessment tool. The design lends itself to statistical analyses that can identify optimal levels on an assessment that best differentiate the simulation groups from the comparison groups (192). In simulation studies, invalid responding base rates are known (e.g., 33% or 50%), and known base rates can be used to generate predictive value statistics tables (see below).

Simulation studies enable researchers to explore novel questions under well-controlled experimental conditions. These studies benefit from group randomization, matching, and knowing which subjects are simulating deficits, ultimately leading to high levels of internal validity. However, the external validity of simulation studies and their generalizability to real-world settings is significantly limited (155; 192; 218). Simulation studies often utilize convenience samples of college undergraduates for their simulation study groups, which may not translate to the performance of actual malingerers in medico-legal settings. The monetary reward in a research study (typically less than $100) pales in comparison to potential gains in a forensic context, as do the consequences of

"being caught" as an exaggerator (192). Researchers can improve the generalizability of simulation study results by recruiting participants from diverse, community-based populations and adding clinical comparison groups (see below) to research designs.

Studies using a known-groups design classify feigning and non-feigning groups *a priori* using an independent standard or criterion (i.e., failure of an RVT), and then systematically analyze the similarities and differences between the "known" groups (192; 217; 218). Known-groups research is thought to have high generalizability, as data are usually collected in "real-world" contexts, such as litigants or disability claimants who are undergoing a neuropsychological evaluation (192). Known-groups designs address major limitations in simulation studies by utilizing data collected in settings where invalid responding is likely to occur and where examinees are likely to be motivated by real-world incentives (218).

The main problem of known-groups designs stems from the reliable and accurate classification of the criterion groups. It is highly unlikely that people in real-world settings with real-world incentives will openly admit after testing that they did, in fact, perform worse than their actual abilities. Researchers must therefore rely on operationally defined "gold standards" such as the 1999 Slick et al. criteria for malingered neurocognitive dysfunction (MND) or the 2005 Bianchini et al. criteria for malingered pain-related disability (MPRD). As with any "gold standard" in a burgeoning scientific field, these proposed criteria have critics (35) and are subject to change based on scientific developments (246). As a result, known-groups designs utilize a "best-available classification" system that may ultimately require retroactive group reclassification and data reanalysis. Known-groups designs are also limited by

researchers losing the ability to randomly assign participants to groups of malingers and non-malingers (155). Additionally, researchers cannot be totally sure of the base rates of invalid performance in their samples, limiting the accuracy of predictive value statistics that can be generated from the data. Consequently, researchers using a known-groups design must rely on base rate estimates to generate their predictive value statistics (183).

Recently, Rogers (218) recommended combining simulation and known-groups designs into a "combined groups" design. This approach would allow researchers to benefit from the internal validity of simulation studies and the generalizability of known-groups designs (192; 218). The additional group allows researchers to verify that between-group differences reflect invalid performance rather than facets of a clinical condition (61). However, adding clinical comparison groups to a simulation study is not without risk to internal validity. The added clinical group would preclude full-study randomization of group assignment, and there is no assurance that the clinical group would be classified correctly (155). Researchers can implement manipulation checks to their simulation study protocol to increase confidence that the groups are responding in the desired manner.

Stevens and Merten (249) recently used a combined groups design to compare reaction time latency and variability between three groups of forensic subjects with and without brain injury and a fourth group of experimental simulators. They found that subjects who failed a freestanding RVT performed significantly worse on cognitive testing, but that healthy simulator performance overlapped considerably with real-world clinical group performance (249). Taken independently, a simulator study would have provided insufficient information to use in a clinical setting, and a known-groups study

using only a pass/fail criterion would not have a "known simulator" comparison group. By combining these research design elements, the authors were able to enhance the internal and external validity of their results. Unfortunately, relatively few researchers have adopted this approach to date, highlighting the need for additional research using the combined groups design (192).

BASE RATES OF INVALID RESPONDING

Base Rates of Invalid Responding in Civilian Populations

Using the aforementioned research designs, researchers have been able to estimate base rates of invalid responding in a variety of settings and contexts. In general, the prevalence of invalid responding varies depending on clinical conditions being evaluated and context of the evaluation (220). Largely in part to a surge of interest and research towards the construct of malingering in the 1990s (254), researchers were able to estimate base rates of malingering or invalid responding in a host of civil and criminal settings (154).

Generally, noncredible performance in clinical assessment becomes more likely in forensic settings and other contexts where outcomes involve the possibility of secondary gain (163; 224). Based on a survey completed by 131 board-certified clinical neuropsychologists experienced in forensic work that encompassed 33,531 neuropsychological evaluations, Mittenberg, Patton, Canyock, and Condit (183) found that 32.7% (95% *CI*: ±4.10) of disability cases, 30.4% (95% *CI*: ±3.64) of personal injury cases, 22.8% (95% *CI*: ±5.83) of criminal cases, and 8% (95% *CI*: ±1.56) of general medical cases involved probable malingering and symptom exaggeration after adjusting for referral source. Mittenberg and colleagues (183) also reported that base rates of

probable malingering (also adjusted for referral source) differed considerably by diagnosis, with the top three being 41.2% (95% *CI*: ±4.51) of mild head injury cases, 38.6% (95% *CI*: ±5.54) of fibromyalgia or chronic fatigue cases, 33.5% (95% *CI*: ±5.50) of pain or somatoform disorder cases

Larrabee (153) combined results from 11 studies published between 1978-2002 with information relevant to malingering base rates in mild traumatic brain injury (mild TBI) litigants. Of the 1,363 subjects, 40% were identified as having performance deficits associated with malingering, ranging from 15% (265) to 64% (115). Despite the wide range of research methodologies used in the 11 studies, Larrabee's (153) results were closely related to the adjusted 41.2% malingering base rate of mild head injury cases reported by Mittenberg and colleagues (183). The similar base rates of probable malingering in compensation-seeking or litigating mild traumatic brain injury cases enhances the confidence of these collective findings.

Two studies conducted by Rogers, Sewell, and Goldstein (222) and Rogers, Salekin, Sewell, Goldstein, and Leonard (221) reported forensic psychologists' estimates of malingering base rates in forensic and nonforensic settings. The 1994 study reported malingering base rates of 15.7% and 7.4% in forensic and nonforensic settings, respectively, and the 1998 study reported base rates of 17.4% and 7.2% in forensic and nonforensic settings, respectively. However, Berry and Schipper (25) contend these data are limited by a lack of well-validated objectives malingering assessment techniques available at the time (i.e., mid-1990s), and propose these relatively lower rates likely represent "floor" rates of psychiatric malingering. Berry and Schipper (25) also argue that successful malingers by definition escape detection and would thus not be included

in the base rates obtained via surveys conducted by Rogers et al. (221; 222) and Mittenberg et al. (183). As such, base rate data obtained from surveys may underestimate the actual base rates of malingering.

Recent studies report a sizable minority of malingering and symptom exaggeration among TBI patients, even in nonforensic or non-litigating settings. Kirkwood and Kirk (146) examined 193 consecutively referred mild TBI patients aged 8-17 and found 17% of them failed the Medical Symptom Validity Test (MSVT; 94), despite no apparent external incentive to perform poorly. Locke, Smigielski, Powell, and Stevens (165) reported a 21.8% failure rate on the Test of Memory Malingering (TOMM; 263) in a sample of 87 consecutively referred, treatment-seeking adult patients with acquired brain injury of all severities. Contrary to Mittenberg et al.'s (183) survey findings, where the observed malingering base rate in mild head injury cases was vastly greater than moderate-to-severe head injury cases, Locke and colleagues (165) found no statistical difference in TOMM failure rates between mild and moderate-to-severe TBI patients.

In Social Security disability evaluations, Chafetz (49) reported adults performed at-or- below-chance levels (i.e., definite malingering per Slick et al. 1999 criteria) 36.5-47.4% of the time on either the TOMM or MSVT, and 45.8-59.7% of claimants failed one or both of the SVTs using their cut score criteria (i.e., probable malingering). As cited by Chafetz (48), Miller and colleagues reported that more than half of their Social Security disability claimant group failed at least one SVT. Other studies of Social Security disability evaluations identified probable or definite malingering between 42-45% (48; 52).

Base Rates of Invalid Responding in Active Duty U.S. Service Member and Veteran Populations

Base rates of invalid neuropsychological assessment performance among active duty U.S. military Service Members and Veterans[3] deserve special consideration, since it is well-established that military service involves health risks that can lead to physical and psychological disability (122; 123). Two military-specific evaluations—medical evaluation boards (MEBs) and compensation and pension (C&P) examinations—are employed to determine the extent of such disability (65; 209). MEBs are comprehensive evaluations that determine whether active duty Service Members are medically fit for duty. If MEBs determine that a service member must be medically separated from duty, the service member is typically given benefits commensurate with the level of disability (209). C&P evaluations are a separate disability evaluation within the Veterans Administration (VA) healthcare system, and they also determine the extent (i.e. percentage) of a Veteran's disability that is related to his or her military service (i.e. service-connected disability percentage or "service connection"). Unlike a MEB, a Veteran can initiate a C&P evaluation at any time beyond the end of his or her military service and potentially receive a service connection. Service connection provides monthly monetary compensation and other ancillary benefits such as tuition assistance for dependents, access to services, and assistance in the home (284).

After more than ten years of warfighting in Iraq and Afghanistan, several studies have explored the prevalence of invalid responding in military-related neuropsychological assessment. A recent survey of 168 psychologists performing

[3] "Veterans" in this manuscript refers to any person who has served in the military at any point and who is eligible for VA benefits. The term "combat Veteran" will be used to describe the smaller subset of Veterans with combat deployment experience.

neuropsychological assessments in VA healthcare system estimated that 42% of Veterans fail RVTs during C&P evaluations and 25% fail RVTs during routine clinical referrals (283). Given that survey data tend to underestimate base rates of a given condition (25), then one might expect the true rates to be higher. A range of studies using various research methodologies to examine invalid responding behavior among active duty Service Members and Veterans have reported RVT failure rates ranging from 17% to 68% (170). Taken at face value, these numbers can be quite alarming. However, reviewing these studies with a critical lens allows a better understanding of how evaluation context and other factors may influence invalid responding base rates in military populations (193).

Armistead-Jehle (8) reported a 58% failure rate on the MSVT (94) in 45 U.S. Veterans referred for clinical evaluation of possible postconcussive symptoms at a Veterans Affairs Medical Center (VAMC). There were no differences in gender, age, education, ethnicity, previous posttraumatic stress disorder (PTSD) or substance use disorder diagnoses between groups of people who passed the MSVT and those who failed it. Additionally, symptom validity scales from the Personality Assessment Inventory (PAI; 186) designed to measure self-reported exaggeration of negative symptoms were not significantly different between groups, suggesting MSVT failure was the result of underperformance in a context heavily associated with secondary financial gain.

Nelson and colleagues (193) examined effort test performance among 119 U.S. Veterans at a Midwestern VAMC. Their sample varied in terms of self-reported concussion history (yes or no), evaluation context (C&P exam or research study), and deployment history (OIF/OEF or non-OIF/OEF). Similar to Armistead-Jehle's (8)

reported 58% MSVT failure rate, Nelson and colleagues (193) reported a 59% (26/44) failure rate on the Victoria Symptom Validity Test (VSVT; 243) among the C&P sample. However, the research sample only had a 10.7% RVT failure rate (8/75). When controlling for effort, the researchers found similar neuropsychological profiles among the Veterans with a history of concussion, supporting previous findings that effort plays a significant role in neuropsychological test outcomes (82; 99).

Drawing from a mixed clinical sample of 286 U.S. Veterans from a VAMC, Axelrod and Schutte (13) reported non-dementia profile patients had a 31.5% (70/222) failure rate on at least one of the MSVT's "easy" subtests (i.e., Immediate Recall [IR], Delayed Recall [DR], and Consistency [CNS]). Of note, only 1% of the overall sample was C&P referrals, while 32% of the sample was referred by mental health providers. This disparity highlights that invalid responding may occur in routine clinical evaluations without direct potential for secondary gain.

Young, Sawyer, Roper, and Baughman (284) examined a sample of 259 Veterans who were referred for neuropsychological assessment at a VA hospital. A total of 74% of the sample were outpatient referrals, 22% were seen for C&P evaluations, and 4% for inpatient hospitalizations. While Veterans with dementia or psychotic disorders were excluded from the sample, 89.6% of the sample was determined to meet criteria for at least one psychiatric diagnosis. The authors reported that 44% of their overall sample (n = 115) failed the WMT (93), with C&P claimants failing the WMT at a much higher rate (71%; n = 41) than clinical outpatient referrals (37%; n = 71). Additionally, the average service connection percentage was significantly higher for those who failed the WMT (*M* = 39.2%, *SD* = 33.8%) than those who passed it (*M* = 27.2%, *SD* = 33.8%).

Russo (226) reported a 68% WMT failure rate among 38 consecutively referred OIF/OEF combat Veterans diagnosed with TBI who presented for follow-up neuropsychological testing in a VAMC setting. 70% of the sample was service connected for disability at the time of the evaluation. Of the 26 combat Veterans who failed the WMT, 44.8% failed all three of the "easy" subtests (IR, DR, and CNS).

In the first study exploring invalid responding among active duty Service Members, Whitney and colleagues (280) reported a 17% MSVT failure rate in a sample of 23 combat Veterans (both active duty and separated from service) from OIF/OEF reporting mild TBI in a VAMC. While their sample consisted of both individuals on active duty at time of testing (n=9) and Veterans no longer on active duty (n=14), all four of the MSVT failures were from the active duty subset. The same four Service Members reported sustaining the mild TBI over five months prior to evaluation with a concussion-related loss of consciousness lasting 10 minutes or less. Three of the four Service Members also failed the TOMM (263), and none met criteria for the MSVT's dementia profile.

Armistead-Jehle and Hansen (10) administered three stand-alone RVTs to a sample of 85 active duty military Service Members that largely consisted of persons reporting a history of mild TBI/concussion (84.7% of the sample) and/or mental health conditions (78.8% had a psychiatric diagnosis). Only seven (8.2%) participants were involved in a MEB process, and none were involved in litigation; as such, the authors concluded that the majority of the participants lacked discernible motivation for secondary gain. Even without a known incentive to do poorly, the overall sample had

failure rates of 20% on the MSVT, 15% on the Nonverbal MSVT (NV-MSVT; 97), and 11% on the TOMM.

While these results appear to be consistent with Whitney et al.'s (280) findings (17% MSVT failure rate), Armistead-Jehle and Hansen (10), considered that their sample's overall results may have underestimated malingering prevalence by including a disproportionate percentage of officers in their sample (54.4%), most of whom were field grade (i.e. middle-to-senior level) officers attending a rigorous, year-long military professional development course (Intermediate Level Education; ILE). Among the non-ILE sample (n = 47), which the authors contended may better represent an active duty military population with respect to age, rank, education, and ethnicity, the authors reported a 30% failure rate on the MSVT (n = 14), a 21% failure rate on the NV-MSVT (n = 10), and a 15% failure rate on the TOMM (n = 7). The non-ILE group's failure rates were higher than the ILE group's 8% failure rates on the MSVT (n = 3), NV-MSVT (n = 3), and the TOMM (n = 3)[4]. The authors argued that underlying variables relating to subgroup membership among active duty and Veteran samples should be closely examined when conducting malingering research in a military population.

Armistead-Jehle and Buican (9) recently conducted the most comprehensive study of performance validity among Service Members on active duty to date. The study's sample consisted of 335 Service Members receiving neuropsychological evaluations at a military TBI clinic, with 117 undergoing a MEB and 218 completing a neuropsychological evaluation for non-MEB/clinical purposes. The authors reported an overall WMT (93) failure rate of 41.8%. The authors also reported that the failure rate of

[4] The same three ILE students each failed the MSVT, NV-MSVT, and the TOMM (P. Armistead-Jehle, personal communication, January 11, 2013).

those undergoing a MEB (63/117; 53.8%) was significantly higher than those undergoing a non-MEB/clinical evaluation (77/218; 35.3%).

In light of the wide range of invalid responding rates in military samples, McCormick and colleagues (170) recently conducted a well-controlled, prospective, multisite study of 214 OIF/OEF combat Veterans (both active duty and separated from service) in "research-only" and "dual[5]" conditions. None of the evaluations were conducted in the context of a C&P evaluation. The authors reported an overall WMT failure rate of 25%, with a 42% (33/78) dual group failure rate and a 15% (21/136) research group failure rate. Failure rates did not differ among those with and without service connected-disability. These results reiterate Nelson and colleagues' (193) findings that invalid responding rates vary in different evaluation contexts.

By combining overall RVT failure rates reported among non-demented and non-psychotic participants in the nine studies from this literature review section (8-10; 13; 170; 193; 226; 280; 284), one would identify 486 RVT failures out of 1340 neuropsychological evaluations, a 36.2% failure rate[6]. However, this figure combines failure rates for disability evaluations, clinical referrals, and research study participants. Among studies that directly reported the data, active duty Service Members and Veterans had a 13.7%[7] RVT failure rate in research-only evaluation contexts (170; 193), but had a

[5] Primarily a clinical evaluation to inform patient care with patient consent for information to be included in research study

[6] 26 failures/45 evaluations using MSVT (Armistead-Jehle, 2010); 34/119 using various RVTs (Nelson et al, 2010); 70/222 using MSVT (Axelrod & Schutte, 2010); 115/259 using WMT (Young et al., 2012); 4/23 on MSVT (Whitney et al., 2009); 26/38 using WMT (Russo, 2012); 54/214 using WMT (McCormick et al., 2013); 17/85 using MSVT (Armistead-Jehle & Hansen, 2011); and 140/335 using WMT (Armistead-Jehle & Buican, 2012).

[7] 8 failures/75 evaluations using various RVTs (Nelson et al., 2010) and 21/136 using WMT (McCormick et al., 2013).

59.3% RVT failure rate[8] in military-related medical disability evaluations (9; 193; 284). It is clear from multiple studies with active duty Service Members and Veterans that invalid responding rates vary depending on evaluation context (9; 170; 193), with the highest base rates among disability evaluations.

Base Rates Summary

As Rogers and colleagues noted in 1993, the prevalence of malingering or invalid responding appears to vary based on context, diagnosis, and population. Perhaps most surprising are the reported base rates from samples who—according to the authors—had no discernible incentive to perform less than their best. Based on research involving adults (165; 193; 221; 222) and children (146) undergoing neuropsychological evaluation without a known external incentive, one can conservatively estimate a 10% base rate of invalid responding on *all* neurocognitive evaluations, regardless of setting or assessment context. As the potential for financial incentive or other secondary gain increases (e.g., disability evaluations, legal claims), the base rate of invalid responding also increases.

Larrabee, Millis, and Meyer (158) contend that, in the presence of potential secondary gain, the likelihood of invalid responding increases to about 40%, plus or minus 10%. Based on the studies reviewed in this section of the manuscript, it is reasonable to consider that range appropriate in civilian cases where secondary gain is most likely and diagnosis of a condition is most subjective. Among U.S. Service Member and Veteran populations, however, invalid responding may occur on as much as 60% of all disability (e.g., secondary gain) evaluations (9; 193; 284). Even in settings

[8] 26/44 using VSVT (Nelson et al., 2010); 41/58 using WMT (Young et al., 2012); and 63/117 using WMT (Armistead-Jehle & Buican, 2012).

where secondary gain was not overtly obvious, it was common to see invalid responding rates range between 8-35% among military populations (9; 10).

In addition to secondary gain, factors such as fatigue, stress, medical conditions, psychiatric conditions, medications, time since injury, initial injury severity, testing environment, examiner skill, assessment instructions, and language/cultural considerations contribute to invalid neuropsychological test performance (11; 181). Collectively, the findings presented in this section underscore the importance of response validity testing in *all* neuropsychological evaluations, even in contexts where secondary gain may not be overt (41; 116).

DIAGNOSTIC VALIDITY AND CLASSIFICATION ACCURACY STATISTICS

The diagnostic validity of a test refers to its ability to differentiate subjects with and without a given condition. Classification accuracy statistics, such as sensitivity (SN), specificity (SP), positive predictive value (PPV), negative predictive value (NPV), and likelihood ratios (LRs), describe diagnostic validity (15; 135; 173). In contrast to group statistics such as *t*-tests or ANOVAs, which classify group differences, classification accuracy statistics are considered to be individual statistics that are useful for determining which subjects are contributing to group differences (157).

Classification accuracy statistics include values that are unique to an instrument and values that rely on the prevalence of a condition. Sensitivity, specificity, and hit rate (HR) indices can be calculated for any instrument without incorporating prevalence of a given condition. Predictive value statistics, on the other hand, calculate values based on base rates of a condition (15). Positive predictive value and negative predictive value,

also known as positive and negative predictive power (PPP; NPP), are examples of predictive value statistics.

To better illustrate classification accuracy statistics, one can refer to the contingency table featured in Table 1. Sensitivity (SN) is defined as the probability of a positive test result in persons who have the condition or characteristic of interest, or the true positive rate. It is the ratio of the number of true positives (TP) to the number of true positives plus false negatives (TP+FN). The formula for sensitivity is:

$$(1) \qquad SN = TP/(TP+FN).$$

Specificity (SP), on the other hand, is the probability of a negative test result in persons who do *not* have the condition or characteristic of interest, or the true negative rate. It is the ratio of the number of true negatives (TN) to the number of true negatives plus false positives (TN+FP). The formula for specificity is:

$$(2) \qquad SP = TN/(TN+FP).$$

Subsequently, the false positive rate can be calculated using the following formula:

$$(3) \qquad FP\ rate = (1 - SP).$$

The hit rate (HR) or overall diagnostic power of a test describes the overall correct classification ability of a measure. It is the ratio of total correct classifications (TP+TN) to total number of subjects evaluated (N). The hit rate is also known as the *efficiency* of the test, or the probability that the test outcome and actual diagnostic condition agree. The formula for hit rate index is:

$$(4) \qquad HR = (TP+TN)/N.$$

The base rate (p) is defined as the prevalence or frequency of a condition of interest in a given population. The formula for base rate is:

$$(5) \qquad p = (TP+FN)/N.$$

As a result of their ratio properties, the values of sensitivity, specificity, hit rate, and base rate range from 0 to 1.00.

Sensitivity and specificity values inversely vary at different diagnostic cutoffs. If a cutoff score is adjusted to increase a test's sensitivity, the test's specificity would decrease, and vice versa. For example, as one changes a cutoff score to correctly identify *more* people who have a disease, one increases the likelihood of making an false positive error (i.e., Type I error).

Receiver operating characteristic (ROC) curves are used to plot the relationship between true positive rates and false positive rates, and can be used to determine the overall accuracy of a test (110; 127; 172; 256). ROC graphs plot a diagnostic tool's true positive rate (i.e., sensitivity) as a function of its false positive rate (i.e., [1-SP]), resulting in a graphical snapshot of classification abilities at varying levels of test outcomes (187). As a result, ROC curves permit researchers and clinicians to determine the optimum cutting score on a psychometric test and display the information on a figure.

The area under the ROC curve (AUC) describes the overall diagnostic power of the test from 0 to 1, where 0 represents a perfectly inaccurate test and 1 indicates a perfectly accurate test (168). By definition, a combined sensitivity and false positive rate of 0.50 represents a 50% chance of making a correct diagnosis. As such, a test with an AUC of 0.5 demonstrates classification accuracy no better than chance. The closer the AUC approaches 1, the greater the likelihood that the test will identify a true positive and not make a false positive error. AUC values from 0.7 to less than 0.8 are considered

acceptable, 0.8 to less than 0.9 are considered *excellent*, and values greater than or equal to 0.9 are considered *outstanding* (125).

As described above, predictive value statistics incorporate base rates of a condition to predict the likelihood of a measure making a correct diagnosis. Positive predictive value (PPV) describes the probability of a given condition being present given a positive test finding. It is the ratio of true positives (TP) to all positive scores (TP+FP), and is expressed in the following formula:

$$PPV = (TP)/(TP+FP). \quad (6)$$

The previous formula assumes that one knows with certainty which results were true positive and which results were false positives. Since this knowledge is rarely known, one can incorporate base rates (p) into the formula:

$$PPV = (p*SN)/[(p*SN)+(1-p)(1-SN)]. \quad (7)$$

Negative predictive value (NPV) describes the probability of a given condition being absent given a negative test finding. It is the ratio of true negatives (TN) to all negative scores (TN+FN), and is expressed in the following formula:

$$NPV = (TN)/(TN+FN). \quad (8)$$

Like PPV, the basic NPV formula assumes that one already knows which persons did and did not have a condition. When the presence or absence of a condition is unknown, one can incorporate base rates (p) into the formula:

$$NPV = [(1-p)SP]/[((1-p)SP)+p(1-SN)]. \quad (9)$$

It should be noted that PPV and NPV vary as a function of base rates; using equivalent test outcomes, PPVs would increase with higher base rates (NPVs would decrease). As base rates decrease, NPVs would increase and PPVs would decrease.

Likelihood ratios (LRs) reflect the percent chance that a person has a condition when testing positive and vice versa (102). The positive likelihood ratio (LR+) is the true positive rate divided by the false positive rate, and is expressed in the following formula:

$$(10) \quad LR+ = SN/(1-SP).$$

Lastly, the negative likelihood ratio (LR-) is the false negative rate divided by the true positive rate, and is expressed in the following formula:

$$(11) \quad LR- = (1-SN)/SP.$$

EXISTING TYPES OF SYMPTOM VALIDITY TESTS

The National Academy of Neuropsychology (NAN) and the American Academy of Clinical Neuropsychology (AACN) call response validity assessment an essential part of any neuropsychological evaluation (41; 116). To accommodate this end, clinicians can utilize tests and indices specifically designed to detect invalid responding. As described earlier, these tests are collectively known as *symptom validity tests (SVTs)*, although they can measure validity in both symptom report and performance.

Symptom validity tests can describe independent, freestanding tests (98; 263) or embedded validity indices derived from existing self-report or neurocognitive measures of attention (198), memory (21), and psychomotor speed (195). Both freestanding and embedded SVTs have unique strengths and weaknesses, and they are generally seen as complimentary tools for symptom validity assessment. Given that an examinee's cognitive effort may fluctuate during the course of long neuropsychological test battery, examiners often use multiple freestanding and embedded SVTs throughout a battery (36). The AACN recently recommended that both freestanding and embedded SVTs be used in

neuropsychology evaluations involving the potential for secondary gain, with embedded measures used at a minimum if time is constrained (116).

Freestanding Symptom Validity Tests

Freestanding SVTs, also known as "stand-alone" SVTs, are designed to detect response bias or exaggeration of deficits while appearing to examinees as a cognitive test (e.g., test of memory, attention, processing speed). Freestanding SVTs are considered to be the most accurate and well-studied type of SVT (27). Though not limited to the format, the most commonly researched freestanding SVTs utilize a forced-choice paradigm (244). The essential characteristic of the forced-choice test is to identify below-chance responding on a series of multiple, two-alternative presentations of words, digits, or patterns. Forced choice tests use the *z* approximation of the binomial theorem (Equation 12 applies when there is a 50% probability of responding correctly) and empirically-derived cutoff scores to identify significantly worse than chance responding or insufficient effort (132). If a subject responds significantly worse than chance based on a certain cutoff score, there is strong evidence for intentionally avoiding the correct answer (85).

$$\text{(12)} \qquad \text{Uncorrected } z \text{ score} = \frac{(\#\text{ of errors}) - 0.5(\#\text{ of items})}{\sqrt{0.25(\#\text{ of items})}}$$

Freestanding SVTs can be categorized by the type of stimuli that form the basis for the test, such as digit recognition tasks, letter- and word-based tasks, and visual or mixed verbal-visual tasks. Many of the earliest freestanding SVTs used digit sequences. One of these digit recognition tasks, the Hiscock and Hiscock (121) Forced-Choice Test, required subjects to choose between two five-digit numbers shown on a card, one of which was seen by the subject prior to a brief delay. This test was also known as the

Hiscock Digit Memory Test (HDMT), and it became one of the first widely used SVTs (105). The HDMT presented a five-digit string of numbers at increasing (5, 10, 15 seconds) lengths of time before asking the participant to choose between two numbers, one of which the subject had seen before. In 1993, Binder created the Portland Digit Recognition Task (PDRT), a visual recognition task of orally-presented, similar five-digit number combinations, classified as "Easy" or "Hard" items. The PDRT includes an interference task of counting backward aloud during intervals, making it more difficult for well-motivated subjects than similar tests without interference (132). The Computerized Assessment of Response Bias (CARB; 1), as its name implies, is a computer-based forced-choice task that presents five-digit number to the examinee for a few seconds. The Victoria Symptom Validity Test (VSVT; 243) is another computerized forced-choice digit recognition test that contains 48 "Easy" or "Hard" items, characterized by the number of shared digits between the target (i.e., correct response) and the foil (i.e., incorrect response similar to the correct response). The VSVT program uses Bayesian analyses and response latencies to identify invalid responding (132).

Another group of freestanding SVTs uses letters or words instead of digits. The Word Memory Test (WMT; 98) and Medical Symptom Validity Test (MSVT; 94) are well-researched word-based tests of cognitive effort. Both WMT and MSVT are computerized effort tests that appear to be verbal memory tests of word pairs. The tests assess immediate forced-choice recognition, delayed forced-choice recognition, consistency of responses, delayed cued recall, and delayed free recall. The MSVT is similar to the WMT, but uses a smaller word list and is faster to administer. Recently, Tombaugh and colleagues (113; 264) developed the Computerized Tests of Information

Processing (CTIP), a computer-based reaction time and processing speed measure. The CTIP uses three reaction time subtests to determine invalid responding: 1) simple reaction time to a repeatedly occurring letter, 2) choice reaction time to a forced-choice word recognition task, and 3) semantic search reaction time to determine whether a word belongs semantically to a given category (163). Other freestanding SVTs using letters and numbers include the Letter Memory Test (LMT; 130), the 21-Item Test (133), and the b Test (38).

Some freestanding SVTs use visual stimuli or a combined visual-verbal format. Swiss psychologist André Rey designed the original Dot Counting Test (DCT) in 1941. In 2002, Boone, Lu, and Herzberg developed a slightly different version of the Rey DCT. Both Rey's and Boone et al.'s Dot Counting Tests examine whether total dot counting time is related to increasing task difficulty. Rey (214) also developed the Fifteen-Item Test (FIT), a non-forced-choice SVT that presents fifteen designs to a subject for a brief period of time and later has the subject reproduce as many designs as possible**.** The FIT is also called the "Rey Memory for 15 Items Test" (132), "Rey's Memory Test" (24), and the Rey 15-Item Memory Test" (232). Although its utility as a symptom validity test is poor (216), the FIT remains one of the most commonly used SVTs (239; 247). The Test of Memory Malingering (TOMM; 263) is a visually-based, forced-choice SVT that asks participants to recognize line drawings presented on either paper booklets or computer monitors. Green's (97) Nonverbal Medical Symptom Validity Test (NV-MSVT) is the visual equivalent of the previously-described MSVT (94), using forced-choice recognition memory of 10 visually-presented color image pairs. Some less frequently used SVTs that employ visual and mixed visual-verbal formats include the Amsterdam

Short-Term Memory Test (ASMT; 229), the Validity Indicator Profile (VIP; 83), and the Coin-in-the-Hand Test (141).

Other freestanding SVTs include self-report measures designed specifically for the identification of symptom exaggeration. The Structured Interview of Reported Symptoms (SIRS; 219) has been described as "the gold standard" for examining malingered mental illness in the field of forensic psychology and psychiatry. The Miller Forensic Assessment of Symptoms Test (MFAST; 180) is also popular among criminal populations in determining competency to stand trial (136). The Structured Inventory of Malingered Symptomatology (SIMS; 281) contains 75 true-false items designed to provide a self-administered malingering screening in about 15 minutes. Designed specifically to evaluate symptom exaggeration in PTSD claimants, the Morel Emotional Numbing Test (MENT; 184) has extensive evidence supporting its ability to identify invalid responding (176; 185).

Embedded Validity Indices

Unlike freestanding symptom validity tests, which are designed with the specific purpose of identifying invalid responding, embedded validity indices are empirically derived from commonly administered neuropsychological or psychiatric tests (234). Embedded validity indices may consist of a single, specially developed test score (e.g., Reliable Digit Span [RDS]; 101), combinations of scores (e.g., Vocabulary minus Digit Span [VDS] on the Wechsler Test of Adult Intelligence-III [WAIS-III; Wechsler, 1997]; 59), or standard clinical scores that have been shown to discriminate between malingerers and valid responders (27). Embedded measures provide examiners several advantages when used in lieu of or in conjunction with freestanding SVTs. First, embedded indices

allow for efficient assessment of performance validity without adding to the time constraints of a typically lengthy assessment battery (27; 36; 234). Next, embedded validity indices appear to be less vulnerable to the effects of coaching than forced-choice SVTs (234). Instead of solely testing one's "memory," as most freestanding SVTs purport to do, embedded measures can be derived from tests across multiple cognitive domains, such as attention, processing speed, executive functioning, and psychomotor speed (36; 234). Adding embedded validity indices to an assessment increases the overall number of SVTs used in a test, increasing the likelihood that negative response bias will be detected (27; 234). Lastly, embedded measures allow performance validity assessment at multiple time points, which is important since a person's responding behavior may vary within and across tests throughout an assessment (36; 116).

Embedded validity indices can also be derived from self-report inventories and questionnaires. Generally, the "hallmark of functional and simulated disorders on these paper-and pencil scales and inventories is abnormally exaggerated complaints—whether in their variety, severity, or both" (163, p. 858). It is common for neuropsychological assessments to include measures of emotional functioning and personality, and many of these measures contain built-in validity scales. The Minnesota Multiphasic Personality Inventory-2 (MMPI-2; 43), for example, includes three validity scales (i.e., L, F, and K) along with its ten clinical scales. Several validity scales aimed at identifying varying forms of symptom exaggeration and noncredible responding have been derived for the MMPI-2, including the Symptom Validity Scale (abbreviated as "FBS" since it was originally called the "Fake Bad Scale"; 162), the Response Bias Scale (RBS; 88), the Henry-Heilbronner Index (HHI; 120), and the Meyers Index (177). The Personality

Assessment Inventory (PAI; 186) contains four measures of response bias and validity, Inconsistency (ICN), Infrequency (INF), Negative Impression (NIM), and Positive Impression (PIM). The Millon Clinical Multiaxial Inventory-III (182) uses three modifier indices to evaluate symptom validity: Disclosure (i.e., how much psychological information one is revealing), Desirability (i.e., under-reporting), and Debasement (i.e., over-reporting). It should be noted that self-report inventories and questionnaires are generally less sensitive to detecting *bona fide* response bias than performance-based validity indices, but nonetheless prove useful in describing both the internal consistency of a patient's complaints and how those complaints align with a patient's cognitive test performance and medical status (163).

Schutte and Axelrod (234) describe several methods used to empirically derive embedded validity indices: considering floor effects in determining variables of interest, incorporating forced choice components into existing tests, and conducting studies using simulation and known-groups (i.e., criterion variable) research designs. Regarding floor effects, one would expect performance by less-injured (i.e., mild TBI) persons on a cognitive test to be generally better than persons with more severe injuries (i.e., moderate-to-severe TBI). Differential prevalence designs can identify indices where lesser-injured persons perform worse than their more severely-injured peers, something that may later be used to identify invalid responding. Tests of learning and memory that include a forced-choice component can be used to identify response bias, such as the California Verbal Learning Test-II (CVLT-II; 63), Rey Auditory Verbal Learning Test (RAVLT; 231), Warrington's Recognition Memory Test (RMT; 274), and Seashore Rhythm Test (236). Cut scores on these forced-choice measures are determined by

statistically comparing performance between known groups (e.g., honest responders, malingerers, clinical patients). Additionally, prospective and retrospective simulation and criterion variable study designs can identify indices with between group differences which can later be run through classification accuracy statistics.

Limitations of Freestanding SVTs and Embedded Validity Indices

Freestanding symptom validity tests and embedded validity indices each have unique limitations that merit specific discussion. A plethora of information about SVTs and how to "beat" them are readily available on the internet, threatening test security (20). Widespread coaching and access to SVT information has led to several well-validated SVTs losing their sensitivity to identify invalid responding in the past few decades (223). Forced-choice paradigms, the most common type of freestanding SVT, are particularly easy to identify and are vulnerable to coaching effects (234). Unscrupulous attorneys or individuals pending medico-legal evaluations can easily describe to others which tests appear to "test one's memory" using digits, words, letters, or visuals while actually testing for invalid responding. Freestanding measures can be time-consuming, a critical factor in neurocognitive assessment where available time and patient tolerance are constrained. Examiners often only have time to administer one freestanding SVT, if any at all (116). If only one invalid responding measure is administered at a single time point in an evaluation, the measure may not adequately characterize the individual's responding behavior throughout the evaluation (36).

Embedded validity indices offer many advantages over freestanding SVTs, such as increased test efficiency, assessment in multiple cognitive domains other than "memory," less vulnerability to coaching, and assessment at multiple time points (234).

However, embedded validity indices are generally inferior to freestanding SVTs in their individual ability to accurately identify invalid responding (179). Examiners must also consider correlations between multiple embedded indices, as multiple failures on essentially similar indices do not necessarily provide convergent evidence of invalid responding (224). Boone (36) recommends using multiple, modestly correlated embedded indices in conjunction with select freestanding SVTs to maximize the likelihood of correctly identifying invalid responding behavior. To limit the effects of shared variance among embedded indices, it may be useful for examiners to use embedded indices from multiple cognitive domains in an assessment.

Recently, Schutte and Axelrod (234) summarized the embedded validity research pertaining to mild TBI, presenting sensitivities and specificities for varying cut scores on embedded indices within tests of attention/processing speed, motor functioning, visuospatial functioning, executive functioning, visuospatial memory, and verbal learning/memory. Not surprisingly, sensitivity and specificity fluctuated as a function of the cut score, where lower cut scores (i.e., lower threshold to "fail") had higher specificity and vice versa. Consistent with research associating invalid responding with slower performance (14), the embedded indices with the best combination of sensitivity and specificity were most commonly derived from tests of reaction time and reaction time variability.

REACTION TIME AND EMBEDDED VALIDITY INDICES IN CONTINUOUS PERFORMANCE TESTS

Reaction time (i.e., response time), well-recognized as a sensitive metric for detecting brain damage, can also be used to detect malingering (112). Reaction time was

commonly used as an indicator of deception in the early 20th century (91; 189). However, the practice largely fell out of favor in the early 1930s, largely replaced by measures of autonomic arousal to detect deception (39).

Recently, computerized cognitive testing has rejuvenated the use of reaction time as a useful measurement of invalid responding (113; 230; 271; 282). In surveying the available literature, simple choice reaction times consistently appear to be slower during invalid versus honest responding (39; 140), suggesting reaction times are delayed when planning and executing an invalid response (282). Reaction time variability, another common cognitive ability measurement, has been studied in patients with brain damage (40), HIV/AIDS (73), and ADHD (194). Willison and Tombaugh (282) recently reported that greater reaction time variability can detect simulated TBI, since the formulating and executing of simulation strategies during testing increases response variability.

Omission and commission errors can be used to measure distractibility and impulsivity, respectively (55). These variables are commonly measured in continuous performance tests, which feature multiple trials appearing in rapid succession over a given length of time. Omission errors occur when a subject fails to respond (e.g., presses a button) within a given time limit, usually the length of the trial. Conversely, commission errors occur when a subject over-responds (e.g., multiple button presses) or incorrectly responds (e.g., presses a button instead of inhibiting a button press) on a given trial. Omission and commission errors have recently been investigated as invalid responding detection metrics (42; 48; 160; 191; 198).

Popular computer-based, continuous performance measures of attention, such as the Test of Variables of Attention (TOVA; 161) and Conners' Continuous Performance

Test-II (CPT-II; 55) measure sustained attention and concentration using multiple variables, including reaction time, reaction time variability, omission errors, and commission error scores. Many of the variables derived from continuous performance tests have been found to be useful in differentiating valid from invalid responding. For example, Leark and colleagues (160) administered the TOVA to a sample of 36 undergraduate volunteers, who either took the test under a "faking bad" instruction set or a "normal conditions" instruction set. After counterbalancing for order of instruction set, the authors reported that the group's "faking bad" responses had significantly more omission and commission errors, slower reaction time mean, and greater reaction time mean variance than the group's "normal conditions" responses. Using a known-groups design and drawing from a sample of 52 neuropsychological evaluation referrals (fifty for mild TBI involved in personal litigation, one for fibromyalgia-related disability, and one for chronic pain-related disability), Henry (118) also reported that TOVA omission and commission errors, reaction time, and reaction time variability were all significantly greater in the "probable malingering" group than the "not malingering" group.

Multiple researchers have reported that omission errors on the CPT-II demonstrate acceptable classification accuracy for invalid responding (42; 198). Lange and colleagues (149) reported CPT-II omissions, commissions, and perseverations may be useful to rule in poor effort but not necessarily rule it out. Ord and colleagues (198) reported that the CPT-II's Hit Reaction Time Standard Error (i.e., reaction time variability) demonstrated acceptable classification accuracy for invalid responding.

Common dependent variables in continuous performance tests—reaction time, reaction time variability, omission errors, and commission errors—appear to serve as

useful embedded validity indices capable of reliably differentiating invalid responding from valid responding. Newer tests, such as the CTIP (113), also demonstrate considerable promise as a freestanding SVT by using reaction time and reaction time variability for detecting invalid responding (211). Unlike traditional forced-choice SVTs that primarily rely on digit, letter, or word recognition and memory, continuous performance tests can assess invalid responding across a variety of cognitive domains, including attention, executive functioning, and processing speed (33; 54). The next section will discuss attentional processes in greater detail to provide insight into why continuous performance tests may be uniquely suited for assessing invalid responding.

ATTENTION AND OCULOMOTOR FUNCTIONING

Conceptually, attention serves as a basic set of mechanisms that facilitates one's awareness of the world and the voluntary regulations of thoughts and emotions (204). Neurological disease and dysfunction can impair attentional processes, making it a prime target for brain disorders research. Aspects of attention can be manipulated and controlled experimentally, providing researchers and clinicians with a window into the underlying neuroanatomical functioning of a patient. Furthermore, attention has been described as a cognitive process sensitive enough to detect impairement from TBI (54).

Vision involves a continuous engagement and disengagement of attention, where individuals fixate their attention on an object, then disengage their attention in order to fixate on a new object. Oculomotor functioning has been said to blend cognition and perception, where eye movements represent one's cognitions, expectations, and motivations for comprehension (74). Visually guided eye movements are regulated by central visuomotor structures, the afferent visual system, and the efferent oculomotor

system, the last of which includes the retina, supplementary eye fields, superior colliculus, lateral geniculate nucleus, frontal eye fields, prefrontal cortex, striate cortex, parietal cortex, basal ganglia, and the brain stem (81). These oculomotor pathways—particularly in the frontal eye fields, supplementary eye fields, prefrontal cortex, and parietal cortex—demonstrate extensive overlap with cognitive processes such as attention, working memory, and learning, suggesting that these systems are functionally interrelated (203). Presumably, if continuous performance tests and other tests of attention can be used to detect invalid responding, then eye movements related to attention may also serve as tools for response validity assessment. The next section will describe the types of eye movements that may be used to measure attentional processes.

Fixations and Saccades

Fixations and saccades are complimentary components of eye movements. Fixations occur when a person focuses on a specific stimulus and stabilizes his or her gaze on it. As such, when one "looks" at a given object, that person is "fixating" on that object. Fixations are controlled by both voluntary and involuntary fixation mechanisms (107; 108). Like the name implies, voluntary fixations occur under the control of the individual who willfully moves his or her eyes onto a given object; these fixations are controlled in bilateral cortical fields in the premotor cortex of the frontal lobes (107; 108). Involuntary fixations, on the other hand, "lock" the eyes onto an object once it has been found, and are controlled by secondary visual areas in the occipital cortex (107; 108). Involuntary and voluntary fixations work hand-in-hand with each other. As a voluntary fixation ends, an involuntary fixation begins.

When one breaks his or her fixation or "lock" on a given point of gaze to fixate upon a new object, he or she is likely making a saccadic movement towards a new point of gaze. Saccades are quick, jerky eye movements that occur *between* fixations during the search for visual targets (17; 139). Regulated by the superior colliculus (17), saccades occur very rapidly, lasting 20-50 ms (268). In a given eye movement—from saccadic initiation to the final, involuntary fixation—the saccadic movement itself encompasses only 10% of the total eye movement duration, while the fixation on a target encompasses the other 90% (107; 108). Saccadic movements are ballistic in nature; once initiated, the speed or direction of a saccade cannot be corrected (139). During a saccadic movement, the brain automatically blocks visual input from being processed (107; 108).

Saccadic Processes and Cognition

From a cognitive neuroscience perspective, saccadic eye movements are influenced by conscious (i.e., deliberate) and unconscious (i.e., automatic) responses to internal and external stimuli. According to Fischer's "three-loop" model (78; 79; 81), three processes occur before a saccade is made: disengagement of visual attention, decision to execute a saccade, and calculation of saccade "metrics" (e.g., direction, amplitude, velocity) needed to reach the target. Several factors influence these saccadic processes.

Before a saccade can be generated, attentional disengagement from an object of focus must occur. The individual or the object may facilitate this disengagement, either from the individual consciously shifting focus away from an object, or by the object disappearing (thus temporarily leaving the individual without an object of focus). When

a fixated-upon object disappears, a brief "gap" occurs before an individual shifts his or her attentional focus elsewhere.

Forced visual disengagement experiments, or "gap paradigms," are commonly used in experiments measuring saccades (68; 144). Gap paradigms compare saccadic performance between "gap" and "overlap" conditions of a measure. In an experimentally manipulated "gap" condition, a fixated-upon object disappears for a brief time (usually around 200ms) before a new object appears in the participant's field of view. These "gap" conditions appear to release subjects' fixations on objects for them, freeing subjects to rapidly redirect their attention towards a new object. Conversely, in "overlap" conditions, a new object appears before the fixated-upon object disappears, forcing the individual to "break" the fixation to generate a saccade towards a new object.

Researchers have hypothesized that gaps enable disinhibition of saccadic movement while overlap conditions inhibit new fixations (128). As such, saccadic latencies are typically shorter (i.e., faster or smaller) in gap conditions and longer (i.e., slower or larger) in overlap conditions (80; 275). This "gap effect" is believed to be moderated by attention and mediated by a "fixation release" component (128).

Cues are sensory stimuli of all types—biological, psychological, and environmental—that influence the decision to generate a saccade and the execution (i.e., calculation) of the saccadic movement. Cues that predict an object's location, also known as predictive cues, help reduce the latency of saccades made towards the target (46). These predictive cues orient an individual towards an area of focus. Cues that distract or incorrectly predict an object's location, on the other hand, increase latencies of saccades towards an object (273). These misleading or invalid cues force one to inhibit

saccadic responses or rapidly identify and correct a saccadic error. The relationship between stimulus cues and saccadic latencies suggests a functional, cognitive relationship between attention, attentional networks, and saccadic eye movements (68; 128; 204). Systems designed to quantify eye movement metrics thus appear to be uniquely suited to measure attentional processing.

Eye Tracking Research of Cognitive Functioning

Eye movement processes like fixations and saccades fall within a relatively narrow range of performance metrics for most individuals, making them highly reliable for comparisons between groups with and without a history of brain injury (74). In recent years, several eye tracking systems have been developed to precisely record and quantify these eye movements. Using high speed cameras and advanced processing equipment, eye tracking systems can measure opticokinetic activity and compute data for one's pupillometry, fixation location and duration, and saccadic latency, velocity, and accuracy (69). Eye tracking systems can measure multiple components of fixations and saccades, enabling comparisons between groups of interest.

In the past decade, cognitive neuroscientists have used advancements in eye tracking technology and novel eye tracking techniques to study attention, response inhibition, working memory, processing speed, and executive function (16; 92; 128; 190; 196; 203). Several studies of neurological injuries and neurodegenerative disorders support the idea that eye movements are closely related to brain functioning (190; 203; 240). Crawford and colleagues (58) used eye tracking equipment to record saccadic eye movements, saccadic inhibitory control, and saccadic errors metrics (e.g., saccadic omissions, commissions, and correction latencies) in young and old groups of delirium

patients and healthy controls. They reported the most reliable oculomotor index of dementia severity was the number of error correction failures, or the lack of corrective saccadic responses to omission or commission errors. Eye movement abnormalities have also been reported in patients with schizophrenia (235), and Parkinson's disease (270). Furthermore, a rapidly growing body of evidence suggests that eye movements and fixations directly correspond to attention and executive functions, two cognitive processes commonly disrupted by TBI (74; 117; 148).

Eye tracking studies have found oculomotor deficits in brain injured individuals long after sustaining the injuries. Most self-reported neurobehavioral symptoms diminish after seven-to-ten days following a single, uncomplicated mild TBI, with "full" recovery normally occurring within three months (171). However, poorer oculomotor performance was detected in groups of mild TBI patients three-to-five months after injury (117), mild and moderate-to-severe TBI patients six months after injury (147), and mild and moderate-to-severe TBI patients more than twelve months after injury (148). As such, oculomotor metrics appear to be sensitive to neuronal injury long after "normal" recovery time, and may serve as useful tools for long-term evaluation of brain injuries.

THE BETHESDA EYE & ATTENTION MEASURE (BEAM)

In 2010, this author and his academic advisor developed the Bethesda Eye & Attention Measure (BEAM), a novel, computer-based eye tracking tool designed to assess cognitive function (18). The BEAM was originally conceptualized as a measure to detect cognitive deficits in the post-acute stage of mild TBI. It was designed as a 12-minute, continuous performance test with a multiple trial format. The BEAM utilizes six pseudorandomly presented trial types, each with unique visual stimuli (i.e., cues) that

were designed to elicit specific cognitive processes of attention and executive function. These cues include white arrows, red arrows, and diamonds which may or may not predict the location of a target circle appearance. Gap conditions and overlap conditions are also interwoven into the four counterbalanced blocks of trials. Saccadic and manual (i.e., button press) information is collected on each trial. Reaction time metrics and omission errors are measured on five non-inhibition trial types, and commission errors are measured on a sixth inhibition trial type.

A feasibility study of the BEAM using 11 subjects without a history of head injury found the BEAM to have excellent internal consistency for manual reaction time (all Cronbach's alpha values > .97) and acceptable-to-excellent internal consistency for saccadic reaction time (all Cronbach's alpha values > .74; overall saccadic reaction time Cronbach's alpha = .94; 18). Despite the small sample size, the BEAM was able to elicit gap, alerting, orienting, and executive effects (see 204) with large effect sizes. The trial design accounted for 79.1% of the variance in manual reaction time and 74.8% of the variance in saccadic reaction time. This author concluded that the BEAM may be a psychometrically sound tool to assess attention, executive function, and processing speed in a relatively short amount of time, and further investigation was merited (18).

A subsequent analysis of BEAM data collected from a follow-on study found saccadic and manual reaction time to be significantly correlated with neuropsychological measures of attention, executive function, and processing speed after controlling for age, education, and gender (19). On the same study, manual reaction time was correlated with self-report measures of depression, traumatic stress, and combat exposure, but saccadic reaction time was not, suggesting BEAM saccadic reaction time may be resistant to

common psychological confounds found in neuropsychological assessment (19). A separate study of BEAM data found that saccadic commission errors were negatively associated with executive functions and working memory after controlling for age and education (201).

While originally designed to identify cognitive deficits associated with mild TBI, the BEAM appears to have potential applications in many other contexts requiring neurocognitive assessment. One such area for exploration is response validity assessment. The BEAM is a continuous performance task, a measure that presents a large number of trials in a short amount of time. As described earlier, continuous performance tests have demonstrated utility for discriminating between groups of valid and invalid responders. Reaction time, reaction time variability, omission errors, and commission errors, each identified as embedded validity indices on the CPT-II (55), TOVA (161), and CTIP (113), are calculated in BEAM output data for both manual and saccadic responses.

The BEAM's oculomotor assessment capabilities potentially offer more sensitive response validity metrics than existing continuous performance test metrics. As described above, oculomotor functioning and the BEAM's saccadic reaction time metrics appear to be resistant to confounding effects of depression and intelligence (19; 117; 201). The BEAM presents a unique opportunity to explore valid and invalid performance in both saccadic and manual responding metrics. Manual responding can be compared with existing continuous performance tests, and oculomotor responding can be used to determine utility above and beyond the manual responding metrics.

SCOPE OF PROJECT

Neuropsychological assessment is commonly used for determining impairment from traumatic brain injury, a widely prevalent injury in civilian and military contexts. Valid symptom report and test performance are essential prerequisites for the accurate interpretation of neuropsychological data. Unfortunately, base rates of invalid responding in civilian and military contexts suggest that symptom exaggeration and underperformance are common in neuropsychological assessment. Many freestanding and embedded validity indicators have been developed and derived to detect invalid responding, but these measures are limited by a variety of factors that dilute their classification accuracy.

This dissertation project evaluated a novel eye-tracking tool, the BEAM, as a method for detecting invalid responding in neurocognitive assessment. This project followed Bianchini, Greve, and Glynn's (28) guidelines for symptom and performance validity research by 1) utilizing a method operationalizing the construct of interest (e.g., noncompliance, malingering, invalid performance), 2) reporting sensitivity, specificity, and predictive power, 3) prioritizing specificity over sensitivity when determining the overall classification rate of invalid performance detection techniques, and 4) considering the purity of the criterion groups (valid controls vs. invalid responders) when estimating a technique's classification accuracy. The study utilized a "combined groups" design that incorporated a well-controlled simulator study and a known-group comparison. The intent of the project was to determine the invalid performance classification accuracy of saccadic and manual BEAM metrics with the goal of identifying useful embedded indices that may be used to detect invalid responding.

SPECIFIC AIMS AND HYPOTHESES

How does the Bethesda Eye & Attention Measure (BEAM) perform in discriminating between valid and invalid responding among healthy persons? How do the BEAM's embedded response validity metrics compare with existing response validity tests (RVTs)? How does valid and invalid responding on the BEAM compare between groups of healthy persons and persons with a history of mild traumatic brain injury? Three specific aims of this project were proposed to answer these research questions. Each aim is supported with several specific, testable hypotheses.

Specific Aim 1: To examine relationships between invalid responding and performance on BEAM metrics.

The first aim of the project was to assess the relationship of BEAM metrics to valid and invalid responding. ROC analyses were used to identify BEAM metrics that best discriminate between valid and invalid responding. Classification accuracy statistics were determined for BEAM metrics with significant differences between groups. Subsequent statistical analyses compared those metrics between groups of valid and invalid responders without a history of TBI. Optimal cut scores were identified for significant BEAM metrics.

Hypothesis 1A: The invalid responding group will demonstrate significantly poorer *compliance with test instructions* than the valid responding group. To test this hypothesis, one variable representing the number of trials invalidated from incorrect initial fixations (i.e., not looking at the center of the screen as instructed) was submitted to ROC analyses.

Hypothesis 1B: The invalid responding group will have significantly slower *reaction time* than the valid responding group. To test this hypothesis, twelve reaction time variables were submitted to ROC analyses. These variables included Saccadic Reaction Time (SacRT) and Manual Reaction Time (ManRT) for the overall measure and five individual trial types.

Hypothesis 1C: The invalid responding group will have significantly greater *reaction time intra-individual variability* than the valid responding group. To test this hypothesis, twelve reaction time variability metrics were submitted to ROC analyses. These variables included Saccadic Reaction Time Intra-Individual Variability (SacRT-IIV) and Manual Reaction Time Variability (ManRT-IIV) for the overall measure and five individual trial types.

Hypothesis 1D: The invalid responding group will have significantly more *commission errors* than the valid responding group. Commission errors were measured as a ratio of commission errors per number of successfully recorded inhibition trials. To test this hypothesis, two variables—Saccadic Commission Error Percentage (SacCom%) and Manual Commission Error Percentage (ManCom%)—were submitted to ROC analyses.

Hypothesis 1E: The invalid responding group will have a significantly more *omission errors* than the valid responding group. Omission errors were measured as a ratio of omission errors per number of successfully recorded non-inhibition trials. To test this hypothesis, two variables—Saccadic Omission Error Percentage (SacOm%) and Manual Omission Error Percentage (ManOm%)—were submitted to ROC analyses.

Specific Aim 2: To compare the invalid responding classification accuracy abilities of BEAM metrics to existing RVTs.

The second aim of the project was to compare the BEAM metrics that have been shown to differentiate valid from invalid responding to existing response validity test metrics. To address the hypotheses under this specific aim, ROC analyses were conducted to identify freestanding and embedded response validity test metrics that best discriminated between the valid and invalid responding groups. Subsequent statistical analyses compared those metrics between the experimental groups. Classification accuracy statistics were determined for freestanding and embedded metrics with significant differences between groups. Optimal cut scores were identified for significant freestanding and embedded response validity metrics. The classification abilities of BEAM metrics were then compared to the embedded and freestanding response validity test metrics.

Hypothesis 2A: The BEAM will provide incremental predictive value above and beyond the classification accuracy of *embedded response validity tests*. To test this hypothesis, the results of the Trail Making Test (TMT; 212), Conners' Continuous Performance Test-Second Edition (CPT-II; 55), and the Digit Span subtest from the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; 278) were submitted to ROC analyses. The classification accuracy of TMT, CPT-II, and the Digit Span variables with sufficient AUC were compared to BEAM metrics using logistic regression.

Hypothesis 2B: The BEAM will provide incremental predictive value above and beyond the classification accuracy of *freestanding response validity tests*. To test this hypothesis, the results of the Victoria Symptom Validity Test (VSVT; 243) and the

Medical Symptom Validity Test (MSVT; 94) were submitted to ROC analyses. The classification accuracy of VSVT and the MSVT variables with sufficient AUC were compared to BEAM metrics using logistic regression.

Hypothesis 2C: The BEAM will provide incremental predictive value above and beyond the classification accuracy of *both embedded and freestanding response validity tests*. All variables with sufficient AUC from embedded indices, freestanding measures, and the BEAM were loaded into a hierarchical logistic regression model to determine the relative contribution of each metric.

Specific Aim 3: To evaluate BEAM and embedded RVT performance between simulator study participants and valid responders with a history of mild traumatic brain injury.

The third aim of the project was to compare the performance of the simulator study's valid and invalid responding groups to research participants with a history of mild TBI. The parent study's TBI cohort was screened in order to exclude participants with moderate-to-severe TBI and participants who demonstrated invalid responding. The remaining group of valid responders with a history of mild TBI was compared to the simulator study groups in order to identify any metrics that may incorrectly classify actual clinical group members as invalid responders. Optimal cut scores were identified for the mild TBI clinical group.

Hypothesis 3A: Of the previously identified optimal BEAM and embedded RVT variables, there will be *no* significant performance differences between valid responders with and without a history of mild TBI. To test this hypothesis, BEAM and embedded

RVT metrics were submitted to between-group comparisons and post-hoc analyses. Classification accuracy statistics were calculated with the clinical group.

Hypothesis 3B: Of the previously identified optimal BEAM and embedded RVT variables, there *will be* significant performance differences between invalid responders and valid responders with a history of mild TBI. To test this hypothesis, BEAM and embedded RVT metrics were submitted to between-group comparisons and post-hoc analyses. Classification accuracy statistics were calculated with the clinical group.

CHAPTER 2: Methods

STUDY DESIGN

This dissertation project's study design utilized a combined groups design to maximize internal and external validity of its results. In a combined groups design, researchers apply results from a simulator study to a group or groups that have met a given criteria for classification (i.e., known-groups; 218). In this dissertation project, results from a prospective simulator study were compared to a clinical group of subjects with a history of mild TBI that met criteria for valid responding.

The core component of this dissertation project was a prospective, experimental simulator study that compared neurocognitive and oculomotor performance between groups of healthy persons[9] with and without an experimental manipulation to perform poorly. Group participation was randomly assigned and blinded to examiners. Between-group comparisons of valid and invalid responders were used to generate classification accuracy statistics for the BEAM and other neurocognitive measures that could be compared to previous research.

To enhance the generalizability of the simulator study's results, data from a "real world" TBI sample drawn from the general population was used for clinical comparisons. This TBI data were collected as part of this project's parent study: "Eye Tracking Indicators of Neurocognitive Status after Traumatic Brain Injury" (Principal Investigator: Mark L. Ettenhofer, Ph.D.). The parent study is a correlational study designed to assess and compare cognitive performance in people with and without a history of traumatic brain injury using BEAM and neurocognitive measures as dependent variables. The

[9] No history of TBI or medical conditions/medications that would impact cognitive functioning.

parent study was approved by the Institutional Review Board (IRB) at Uniformed Services University of the Health Sciences (USUHS; see Appendix C: Administrative Documents).

Several neurocognitive measures with empirically derived embedded response validity indices were used to separate the TBI cohort into known groups of valid and invalid responders. The majority of the subjects in the TBI cohort had a history of mild traumatic brain injuries, and most of these subjects met criteria for valid responding. There were insufficient numbers of invalid responders with a history of mild TBI or subjects with moderate-to-severe TBI to power analyses. Given the available data, only subjects with a history of mild TBI who met criteria for valid responding were included in the "known" clinical comparison group. Results from the prospective simulator study were compared to the parent study data.

PARTICIPANTS

As stated above, the dissertation project evaluated three groups of responders: an invalid responding, biased group of responders without a history of TBI (the "BR" group); a valid responding, unbiased group of responders without a history of TBI (the "UR" group); and a valid, unbiased group of responders with a history of mild TBI (the "UR-mTBI" group). The following sections describe the inclusion and exclusion criteria for each of the three groups. Of note, the UR-mTBI group's inclusion and exclusion criteria are equivalent to the ongoing parent study.

BR and UR Groups

The biased responding (BR) and unbiased responding (UR) groups consisted of persons recruited exclusively for this dissertation project. Participants for the BR and UR groups were recruited using flyers, internet advertisements, and hand-outs. Participants were compensated $30 for their involvement in the study unless they were active duty U.S. military or federal employees. The following inclusion criteria were used to screen potential participants: must be 18 years or older, must have fluency or literacy in English (per self-report), must be willing and able to provide informed consent, and must have obtained written permission from supervisor and/or brigade commander if they are a federal civilian or active duty U.S. military. Participants were not allowed to participate in this study if they have ever sustained a traumatic brain injury of *any* severity throughout their lifetime, including any head injuries that involved an alteration of consciousness (AOC). Participants were also excluded if they had a medical condition (e.g., thyroid disorder, sickle cell anemia) or were actively taking medication that could impair their cognitive abilities, if they had any visual impairment that could not be corrected by glasses/contacts, or if they had motor impairment or amputation of one or both upper extremities. Any participant in the UR group that exceeded cut score thresholds on one or more freestanding RVTs and/or two or more embedded RVTs (see Appendix A: Table 2) was excluded from all analyses (i.e., he or she was not analyzed as a BR group member).

UR-mTBI Group

The unbiased responders with self-reported history of mild traumatic brain injury (UR-mTBI) group consisted of participants in this project's parent study who reported of

history of at least one mild TBI. In accordance with DoD/VA clinical practice guidelines (66), mild TBIs were defined as events that involved a sudden movement or a blow to the head the resulted in a loss of consciousness (LOC) ranging from 0 to 30 minutes or loss of memory (i.e. post-traumatic amnesia; PTA) less than or equal to 24 hours. While the American Congress of Rehabilitation Medicine (ACRM) also qualifies alterations of consciousness (AOC) as "mild TBIs" (142), any subject reporting AOC without LOC or PTA was not included in the parent study's (or this project's) mild TBI group. Consistent with other studies using TBI samples (159; 260), head injury information from the parent study sample could not be verified by medical record. To obtain head injury details, examiners used a semi-structured interview that obtained detailed information about injury characteristics, mechanism of injury, and injury sequelae. Follow-up questions were asked as needed to provide a comprehensive understanding of the injury or injuries. A week after the participant completed the assessment, a team consisting of two licensed psychologists with post-doctoral fellowship training in clinical neuropsychology and three-to-five clinical psychology doctoral students carefully considered the accuracy and context of the self-reported injury characteristics and classified the individual based on the person's most severe head injury. The potential classifications included "no TBI," "possible mild TBI (AOC only)," "mild TBI," "moderate TBI," and "severe TBI."

Participants in the UR-mTBI groups were recruited using flyers, internet advertisements, hand-outs, and newspaper advertisements. All UR-mTBI group participants were told prior to the assessment that they would be compensated $40 for their involvement in the study unless they were ineligible for compensation (i.e., active duty military or federal employees). The following inclusion criteria were applied to all

parent study participants: must be 18 years or older, must have a history of one or more head injuries with a loss of consciousness or memory, must have fluency or literacy in English (per self-report), must be willing and able to provide informed consent, and must have obtained written permission from supervisor and/or brigade commander if they were a federal civilian or U.S. military. Participants were excluded from the parent study if they had a medical condition (e.g., thyroid disorder, sickle cell anemia) or were actively taking medication that could impair their cognitive abilities, if they had any visual impairment that could not be corrected by glasses/contacts, or if they had motor impairment or amputation of one or both upper extremities. Parent study participants were excluded from this study's UR-mTBI group if their head injuries were in the moderate-to-severe range (LOC > 30 minutes or PTA > 1 day) or if their head injuries did not involve a loss of consciousness or post-traumatic amnesia (i.e., possible mild TBI [AOC only]). Lastly, parent study participants were excluded from the UR-mTBI group if they met criteria for invalid responding, which was defined as exceeding one or more of the eight empirically derived embedded RVT cutoff score thresholds of 90% specificity or greater (see Appendix A: Table 2).

SETTING AND EQUIPMENT

Setting

The room that used for testing was located in Dr. Mark Ettenhofer's research laboratory (Room B1032 on the USUHS campus). During computer-based tasks, the participant sat at a desk with a computer monitor and eye tracker, and the examiner sat five feet behind the participant at a desk facing 90 degrees from where the participant was facing. During non-computer-based measures, including the semi-structured

interview, the participant and examiner faced each other and used the examiner's desk as the assessment surface.

Computers

Two computers were used during the study: a "stimulus computer" used by participants and a "control computer" used by examiners (see Appendix C: Pictures). The stimulus computer was used to present computerized measures to the subject. The stimulus computer was a Dell Precision T1500 with an Intel Core i7 860 CPU, 2.80 GHz processing speed. Subjects viewed computerized measures on a 15" Asus VW193 flat-screen monitor set to 1440 x 900 pixel resolution. Examiners used the control computer to run programs on the stimulus computer and record eye-tracking data. The control computer was a custom-built PC with a Pentium Dual-Core E5400 CPU, 2.70 GHz processing speed.

Eye-Tracking Device

Eye tracking was performed using an Applied Science Laboratories (ASL) D6 High-Speed (HS) Desktop Eye Tracker (see Appendix C: Pictures). The primary components of the eye tracker included a high speed camera to record visual information from the eye and an infrared illuminator to provide a corneal reflection from which eye gaze vectors can be computed. This infrared illuminator operated within the spectral range of between 760 and 1400 nanometers at intensities of <0.5 mW/cm² to 0.7 mW/cm², well below the maximum safe chronic ocular exposure value of 10 mW/cm². This system used non-coherent illumination; there were no lasers in the system. The

desktop-mounted D6-HS system did not require chinrests or other head stabilizers; it was chosen for its inconspicuous design and for its enhanced participant comfort.

Response Pad

A Cedrus RB-530 response pad was used to record manual (i.e., button press) response time with 1 millisecond time resolution (see Appendix C: Pictures). The Cedrus response pad was chosen to allow a higher level of time resolution relative to buttons on a standard computer keyboard.

Software

ASL Results Version 1.0 was used to analyze eye tracking data. E-Prime 2.0 software, a suite of applications used in computerized experiment design, data collection, and analysis, was used to run the BEAM. E-Prime 2.0 software enabled paradigm developers to use signal codes called "XDATs" to mark events that occur throughout computer-based measures. By marking certain events (e.g., trial begins, target appears, button is pressed, etc.), developers could synchronize participant responses with paradigm activity. SPSS Version 20 was used for statistical analyses.

Data Acquisition and Post-Processing

Data Acquisition Procedure

The D6-HS system used a two-computer interface. Participants completed computerized assessments at the stimulus computer, where the eye tracker and response pad recorded oculomotor activity and manual responses (i.e., button presses), respectively. Examiners sat at a control computer with live-feed video monitors and an ASL data processing unit. Cables connected the two computers, synchronizing

participant responses, eye movements, and assessment events, enabling examiners to calibrate participants and monitor their gaze in real time.

Participants sat with their head approximately 24” from and level with the center of the stimulus computer monitor. The D6-HS was positioned directly below the monitor, facing the participant. The examiner sat behind the participant at the control computer. The examiner oriented the D6-HS camera onto the participant’s right eye. Next, the eye tracker was calibrated by having the participant gaze sequentially at a series of dots presented on the stimulus display. The calibration process took approximately 2 minutes to complete. Data were then be collected at 120Hz by recording eye tracking data synchronized with event markers related to the presentation of stimuli.

A parallel cable connecting the stimulus computer and control computer enabled BEAM events (e.g., trial beginning, stimulus appearing, etc.) to be synchronized in real time with manual and oculomotor data collection. In a given trial, the stimulus computer sent XDAT codes that signaled when trials began, when visual stimuli were presented on screen, when buttons were pressed, and when trials ended. Because every data segment collected during the BEAM uses a specific XDAT code, ASL software was able to perform trial-by-trial analysis after the participant completed the BEAM. The data output enabled examiners to observe BEAM activity during a given trial, identify where a person was looking throughout a given trial, and collect button press data.

Data Post-Processing

Eye tracking data noted above was first be filtered to remove blinks, out-of-range values, and other potential sources of error. Fixations and saccades were then computed with the ASL eye tracking analysis software using established algorithms. Custom

scoring software was then used to perform data acquisition checks to enhance confidence in the obtained data, identifying trials in which momentary visual signal loss prevented reliable reaction time or inhibition error calculations. After screening the data for unsuccessfully recorded trials, the custom scoring software derived task- and trial-dependent variables from eye gaze and motor responses. These values were then collapsed across multiple task trials in order to obtain summary metrics relevant to each task (e.g., median saccadic reaction times for specific trial types, commission error percentages). At least 10 successfully recorded trials (out of 32) were required to obtain a summary median reaction time or inhibition error metric for each trial type.These derived summary metrics were then used in primary analyses of interest, similar to the summary scores of traditional cognitive tests.

INDEPENDENT VARIABLES

There are two independent variables in this dissertation project, each with two levels. The first independent variable is "invalid responding bias" (yes or no). The second independent variable is "head injury" (yes or no). This project included three groups: unbiased responders with a history of mild traumatic brain injury (UR-mTBI), unbiased responders without a history of head injury (UR), and biased responders without a history of head injury (BR).

The head injury independent variable was manipulated quasi-experimentally through recruitment, semi-structured interview, and clinical panel consensus. Invalid responding bias was manipulated in two different ways, depending on the study group. For the parent study's mild TBI cohort, existing embedded RVT cut scores (see Appendix A: Table 2) were used to exclude potential invalid responders. As such,

invalid responding bias in the TBI cohort was manipulated quasi-experimentally. By contrast, invalid responding bias in the non-TBI groups (i.e., groups from the simulator study) were manipulated experimentally through random group assignment and group-specific scenarios presented to participants prior to their assessment. Specifically, simulator study participants were randomly assigned to either BR or UR groups and given a scenario that asks them to perform as if they sustained a head injury (i.e., biased responding; BR group) or to perform their best (i.e., unbiased responding; UR group). To enhance internal validity of the experimental manipulation, examiners were blinded to group assignment throughout the assessment, scoring, and data entry of simulator study participants.

DEPENDENT VARIABLES (MEASURES)

This study compared performance between groups of valid and invalid responders on the Bethesda Eye & Attention Measure (BEAM), neurocognitive tasks with embedded response validity indices, and freestanding response validity tests (RVTs). A well-controlled simulator study incorporating several RVTs into its design allowed the relative sensitivity of the embedded and freestanding RVTs to be calculated and compared to each other (96). Furthermore, by incorporating neurocognitive measures into the design, this study was able to assess the ability of various RVTs to predict whether or not neurocognitive test scores from clinical comparison groups were accurate (96). Sample characterization measures were administered to determine group demographics and identify group differences in age, gender, years of education, race/ethnicity, premorbid intelligence, and knowledge of TBI sequelae.

Sample Characterization Measures

Baseline Interview

Two baseline interviews were used for this study. The first baseline interview was used solely for the simulator study of this dissertation project. The simulator study Baseline Interview obtained demographic information (e.g., age, race/ethnicity), military history (if applicable), educational background, languages spoken, employment/disability status, medical history, medications, and alcohol/nicotine/caffeine use. Please see Appendix C: Simulator Study Baseline Interview. The second baseline interview used in this dissertation project was drawn from the parent study's archival data; this Parent Study TBI Cohort Baseline Interview was used with participants with a self-reported history of TBI. The parent study Baseline Interview obtained similar information from simulator study Baseline Interview plus information related to head injuries, activities of daily living, and treatment history. Please see Appendix C: Parent Study TBI Cohort Baseline Interview.

Feedback Interview

For the simulator study, a lab member other than the examiner administered a post-assessment interview to assess qualitative and quantitative information about test-taking strategies, perceived performance, and experiences of study participation. The feedback interview also assessed the examiner's beliefs towards the participant's group membership. The interview also served as a manipulation check on the primary independent variable of group assignment. Please see Appendix C: Simulator Study Feedback Interview.

Wechsler Test of Adult Reading (WTAR)

The WTAR estimates premorbid intellectual functioning (124). Participants are given a list of 50 words and asked to pronounce the words as best they could. The measure ends when participants reach the 50 word limit or when the participants incorrectly pronounce 12 words in a row. The WTAR's internal consistency (.90-.97) and test-retest reliability (.90-.94) are excellent. The WTAR also positively correlates with Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; 276) Full Scale IQ (FSIQ; .63-.80) and the Verbal Comprehension Index (VCI; .61-.80; 124).

Head Injury Knowledge Scale (HIKS)

The HIKS assesses the level of misconceptions of the effects of brain injury via 18 true/false responses, with the number and type ("minimization" or "overgeneralization") of inaccurate responses reflecting the magnitude and type of misconceptions (197). The HIKS was used to identify any differences in head injury sequelae knowledge between groups that may bias invalid responding approaches (71). For example, if the BR group knew significantly more about cognitive and behavioral sequelae of TBI than the UR group, the BR group may demonstrate more sophisticated invalid responding than what would be expected from the general population.

The HIKS assesses misconceptions across several domains impacted by traumatic brain injury, including physical, sensory-perceptual, cognitive, and behavioral domains. Participants with (Version A) and without (Version B) a history of head injury are asked to indicate whether they think the changes referred to in each item are true ("often or most of the time"), or false ("never or rarely"). The HIKS contains an 8-item

"overgeneralization" subscale and a 6-item "minimization" subscale. HIKS Version B was used in this dissertation project's simulator study (see Appendix C: HIKS B).

Oculomotor Measure

Bethesda Eye & Attention Measure (BEAM)

The BEAM is a computerized, continuous performance task that assesses saccadic (i.e., visual) and manual (i.e., button press) responses to stimuli presented on a computer monitor. The BEAM consists of one block of 24 practice trials and four blocks of 48 trials. Each block is counterbalanced, with equal numbers of trial types (Directional Cue [DC], Nondirectional Cue [NDC], Misdirectional Cue [MDC], Uncued with Gap [UC-G], Uncued with Overlap [UC], and Directional Cue-Red Arrow [DCR]), target locations (up, down, left, and right), and arrow cue locations (up, down, left, and right). For DC, NDC, MDC, UC-G, and UC trials, participants are asked to look at a fixation cross at the center of the screen until a target circle appears above, below, left, or right of the screen's center. Participants are asked to look at the target circle and press a button as soon as a target circle appears. Saccadic and manual reaction time, reaction time intra-individual variability, and omission errors are calculated for the five "non-inhibition" trial types and the overall measure. Reaction time (RT) is represented by a median score, and reaction time intra-individual variability (RT-IIV) is represented by the standard deviation of the reaction times. Overall saccadic and manual RT is calculated by averaging the median reaction times across the non-inhibition trial types. Overall saccadic and manual RT-IIV is calculated by averaging the RT-IIV values from the non-inhibition trial types. Omission errors occur when a participant fails to look at a target circle or press the button by the time a new trial begins (1000ms). On DCR (i.e., inhibition) trials, participants are

told *not* to look at the target circle or press a button when the target circle appears. Unlike the five other trial types, which use only white arrow or diamond cues, DCR trials use red arrow cues. Commission errors are calculated when participants look at the target circle or press the button during DCR trials. Preliminary analyses of BEAM psychometric properties, convergent validity, and divergent validity are discussed above in the literature review.

Of note, built-in data acquisition validity indices enhance the confidence in the obtained reaction time, reaction time intra-individual variability, omission errors, and commission errors variables. Before any BEAM metrics are calculated, a custom-made scoring program checks for lost or missing saccadic data segments. A trial is discarded if the eye tracker loses its lock on a person's pupil or corneal reflection on more than 20% of the segments recorded after a target circle appears. If the trial is not discarded for data loss, it is said to have been "successfully recorded." Next, the BEAM checks to see if participants were following instructions on a trial. Trials are discarded if a person was looking outside the center of screen when a target circle appears (i.e., "Invalid Initial Fixations"). If a trial was successfully recorded and the initial fixation at time of target circle onset was in the center of the screen, the scoring program would calculate BEAM metrics.

Cognitive Performance Tasks with Embedded Response Validity Indices

Conners' Continuous Performance Test, Second Edition (CPT-II)

The CPT-II is a computerized vigilance test that measures attention problems (55). The CPT-II requires examinees to press the space bar as quickly as possible whenever a target (i.e., any letter other than "X") appears on the computer screen, and to

inhibit this response when the letter "X" appears on the screen. Ninety percent of letters presented are targets. Each letter is presented for 250ms, with varying interstimulus intervals (ISIs) of one, two, or four seconds between letters. Full test administration includes a one-minute practice block and six long blocks of trials, with each long block containing three sub-blocks of 20 trials. Overall, the measure takes approximately 14 minutes to administer.

The CPT-II generates 12 indices of responding, including Hit Reaction Time (Hit RT), Hit Reaction Time Standard Error (Hit RT SE), Omissions, Commissions, Variability of Standard Error, Hit Reaction Time Block Change (Hit RT Block Change), Hit Standard Error Block Change (Hit SE Block Change), Hit Reaction Time Interstimulus Interval Change (Hit RT ISI Change), and Hit Standard Error Interstimulus Interval Change (Hit SE ISI Change), Detectability (*d'*)[10], Response Style (β), and Perseverations (55). Eight indices measure inattention: Omissions, Commissions, Hit RT, Hit RT SE, Hit RT ISI Change, Hit SE ISI Change, Variability, and Detectability. The Commission Index and Hit RT also measure impulsivity, along with Perseverations. Hit RT Block Change and Hit SE Block Change measure vigilance and alertness (55; 207).

Each index is designed to measure attention uniquely (55; 207). Hit RT measures the average speed of correct responses for the entire test, and Hit RT SE measures response speed erraticness, with higher scores suggesting inconsistent responding. Omission and Commission Indices identify failures to respond to targets and responses to non-targets, respectively. Variability measures reaction time variability across 18 segments of the test in relation to overall Hit RT SE. Hit RT Block Change describes

[10] The Detectability index is also referred to as the Attentiveness index

changes in reaction time across the duration of the test, with higher scores suggesting a slowing of reaction time. Hit SE Block Change measures changes in response consistency over the course of the test, with higher scores suggesting a loss of consistency. Hit RT ISI Change measures changes in reaction time at different ISIs (i.e., one, two, or four seconds), and Hit SE ISI Change describes the change in reaction time consistency across different ISIs. Detectability measures the examinee's ability to distinguish a target from a non-target. Response Style describes an examinee's responding trends, with higher scores suggesting cautious, accurate responding and lower scores suggesting attempts to respond to all targets in spite of accuracy. Lastly, Perseverations indicate the number of reaction times less than 100ms, reaction times that suggest the examinee is anticipating a stimulus rather than reacting to one.

Chen and colleagues (53) reported that Omissions, Commissions, Hit RT, Hit RT SE, and Variability display acceptable-to-excellent test-retest reliability, ranging from .70-.90. Using a normative population, the test-retest reliability ranged from .55-.84 for the same five measures (55). While the CPT-II has previously demonstrated sensitivity to mild TBI in the chronic phase of recovery (147), CPT-II performance does not appear to significantly differ between TBI severity (i.e. mild, moderate, or severe) in civilian and military samples (149).

According to the CPT-II manual, Response Style, Omissions, and Perseverations can be used to detect invalid responding (55; 207). Response Style T-scores below 40 or greater than 60 suggest overly impulsive or overly cautious responding, respectively. Extremely high T-scores ($T > 100$) on Omission and Perseverations suggest misunderstanding of instructions and inaccurate results. Recently, several researchers

have evaluated the CPT-II using known-groups of valid vs. invalid responders (42; 149; 198). After submitting all CPT-II variables to receiver operator characteristics (ROC) curves, Ord and colleagues (198) reported Omissions (area under the curve [AUC]=0.77, 95% *CI*: .65-.89) and Hit RT SE (AUC=0.77, 95% *CI*: .65-.90) produced the best classification accuracy scores for probable or definite malingered neurocognitive dysfunction (MND; 244) in a sample of 88 all-severity TBI cases with high external incentives. Using Ord et al.'s (2010) data, Schutte and Axelrod (234) reported >19 raw Omissions rendered specificity of 90% and sensitivity of 41%, and raw Hit RT SE values >13 rendered specificity of 90% and sensitivity of 52%.

Busse and Whiteside (42) examined 413 consecutively referred neuropsychological evaluations, and also reported CPT-II Omissions had acceptable invalid responding classification accuracy (AUC=0.76). Using a cut score of >12 raw omissions, the authors reported specificity of 88% and sensitivity of 52% (42). In a separate study of 158 deployed U.S. Service Members with a history of deployment-related mild and severe TBI, Lange and colleagues (149) reported that Omissions (.69 < AUC < .75), Commissions (.76 < AUC < .79, and Perseverations (.70 < AUC < .79)[11] demonstrated the best invalid responding classification accuracy among the CPT-II variables. When comparing the groups of mild TBI participants who either passed or failed effort testing, the authors reported >11 raw omissions had specificity of 91% and sensitivity of 31%, >21 raw commissions had specificity of 86% and sensitivity of 45%, and >1 raw perseveration had specificity of 93% and sensitivity of 43% (149). Similar results were obtained when comparing groups of mild TBI participants who failed effort

[11] Ranges include point AUC results for both T-scores and raw scores among mild TBI-fail vs. mild TBI-pass comparisons and mild TBI-fail vs. severe TBI-pass comparisons.

testing with severe TBI participants who passed effort testing; >11 raw omissions had specificity of 93% and sensitivity of 31%, >21 raw commissions had specificity of 93% and sensitivity of 45%, and >1 raw perseveration had specificity of 90% and sensitivity of 43% (149). Overall, embedded response validity indices in the CPT-II appear to demonstrate some utility towards detecting invalid responding, although the indices may better be used to "rule in" invalid responding rather than ruling it out (42; 149; 198).

WAIS-IV Digit Span Subtest

The Digit Span subtest of the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV; 278) measures attention and working memory, and consists of three separate components: Digit Span Forward, Digit Span Backward, and Digit Span Sequencing. In Digit Span Forward, participants repeat numbers that are spoken to them (i.e., the correct response to "1-2-3-4" is "1-2-3-4"). In Digit Span Backwards, participants repeat the numbers that are spoken to them in the reverse order (i.e., the correct response to "1-2-3-4" is "4-3-2-1"). Lastly, Digit Span Sequencing requires participants to order the numbers that are spoken to them from lowest to highest (i.e., the correct response to "3-2-4-1" is "1-2-3-4"). On each component of the Digit Span subtest, participants gradually proceed with longer and longer digit spans until they incorrectly respond to two digit sequences that span the same length. The Digit Span subtest demonstrates excellent internal consistency (.93) and good test-retest reliability (.82; 278).

The Digit Span subtest has spawned numerous studies examining embedded response validity indices (234; 252; 255). One of the most researched embedded response validity tests (RVTs) in Digit Span is *Reliable Digit Span* (RDS; 101; pp. 219-220), which is "calculated by summing the longest string of digits repeated without error

over two trials under both forward and backward conditions." For example, a participant who correctly responds on both trials with *four* digits forward, correctly responds on both trials with *three* digits backward, but incorrectly responds to one or both trials beyond that point would earn an RDS score of *seven*. Larrabee (153) found RDS scores <8 have specificity of 94% and sensitivity of 50%. Babikian and colleagues (14) reported RDS scores <7 have specificity of 93% and sensitivity of 45%. Recently, Schroeder and colleagues (2012) conducted a systematic review and cross-validation study of RDS, and concluded that RDS can be used effectively in many clinical samples, with cutoff scores <7 having global specificity and sensitivity rates of 96% and 30%, respectively[12].

The age-corrected scaled score (ACSS) on the Digit Span subtest has also been researched as an embedded response validity index. In addition to their RDS findings, Babikian and colleagues (14) found ACSSs <6 have specificity of 93% and sensitivity of 42%. By comparison, Axelrod and colleagues (12) reported ACSSs <6 have specificity of 97% and sensitivity of 36%.

A recent meta-analysis of 24 studies using RDS or ACSS to detect invalid responding found both indices effectively discriminated between valid and invalid responders, with an average RDS Cohen's *d* effect size of 1.34 (95% *CI*: 1.18-1.50) and an average ACSS effect size of 1.08 (95% CI: 1.01-1.50; 137). The same study found both indices demonstrated strong overall specificity (RDS *M* = 86.1%; ACSS *M* = 86.5%) and good sensitivity (RDS *M* = 63.3%; ACSS *M* = 59.7%), with no significant classification accuracy differences between RDS and ACSS (137). A recent study evaluating RDS and ACSS in a pediatric sample (ages 8-16) with the Wechsler Intelligence Scale for Children-4th Edition (WISC; 277) reported ACSS cut scores <6 had

[12] Global specificity and sensitivity rates were calculated using weighted averages.

specificity of 96% and sensitivity of 51%, and RDS cut scores <7 had specificity of 92% and sensitivity of 51% (145). Additionally, depression of all severities and subtypes does not impact RDS or ACSS in the Digit Span subtest (90).

Most of the available studies examining RDS and ACSS have used Digit Span subtests from WAIS-IV predecessors that only included Digit Span Forward and Digit Span Backward components (137; 233). The WAIS-IV, however, has an additional sequencing component to its Digit Span subtest (278), prompting researchers to consider a revised RDS index that includes all three Digit Span subtests. As cited in Young et al. (284), Spencer and colleagues were the first to study the Reliable Digit Span-Revised (RDS-R), which at cutoff scores <12 demonstrated specificity of 89% and sensitivity of 59%. Seeking to further examine the classification accuracy of the RDS-R index on the WAIS-IV Digit Span subtest, Young and colleagues (284) found RDS-R scores <11 demonstrated specificity of 89% and sensitivity of 32%, and scores <12 rendered specificity of 78% and sensitivity of 48%[13]. Reese, Suhr, and Riddle (210) reported RDS-R (which they called "Enhanced RDS") values <12 rendered specificity of 94% in their head-injured sample and 59% sensitivity. Reese and colleagues (210) also studied Alternative RDS (A-RDS), which was calculated by summing reliable digit forward and reliable digit sequencing; they reported ARDS values <10 had specificity of 87% in their head injured sample and 78% sensitivity.

Taken together, the Digit Span subtest appears to have a wealth of research suggesting its embedded response validity indices (e.g., RDS, ACSS) can be useful for detecting invalid responding. Additional indices particular to the WAIS-IV (i.e., ARDS,

[13] These specificity and sensitivity rates are poorer than the classification accuracy statistics cited in the Spencer et al. study.

RDS-R) demonstrate potential utility towards identifying invalid responding. However, more evidence is needed before using RDS-R or ARDS instead of RDS or ACSS.

Trail Making Test (TMT)

The TMT is a graphomotor test consisting of two components, Part A and Part B (212). In TMT A, participants draw lines to connect numbered circles in order; the task largely depends on the participant's psychomotor speed and visual search abilities. In TMT B, participants draw lines to connect circles with alternating numbers and letters; this task places additional demands on the participant's working memory, cognitive flexibility, and executive functioning. The score on each part of the TMT is determined by the time required to complete each trial. The TMT displays sufficient test-retest reliability on both Parts A & B (.79 and .89, respectively; 67), and a moderate-to-sufficient construct validity (.36-.93; 70). Recently, Tsirka and colleagues (267) reported differences between a control group and a mild TBI group on the TMT.

The TMT was one of the first measures to be evaluated for embedded response validity indices (89; 115; 265). Error rates (225), completion times (134), and completion time ratios (89) are some of the most commonly studied embedded indices on the TMT. According to Suhr and Barrash (252), decades of embedded response validity research using the TMT have produced equivocal results. Iverson and colleagues (134) found that TMT A completion times >62 seconds had 100% specificity but only 17% sensitivity to invalid responding; the authors also reported TMT B completion times >199 seconds or more had 100% specificity but only 7% sensitivity. Recently, Busse and Whiteside (42) reported that TMT B was able to distinguish between biased (i.e., invalid) and unbiased (i.e., valid) responders (AUC=.75), with TMT B completion times >119

seconds rendering specificity of 85% and sensitivity of 61%. Drawing from a sample of 76 consecutive mixed acquired brain injury patients being evaluated for outpatient brain injury rehabilitation, Powell, Locke, and Smigielski (205) reported TMT A completion times greater than 47 seconds had specificity of 83% and sensitivity of 72%. Powell and colleagues (205) also reported TMT B completion times >124 seconds had a specificity of 81% and sensitivity of 50%, In summary, slower completion times on the TMT appear to be associated with invalid responding, but embedded TMT validity indices do not by themselves appear sensitive enough to detect invalid responding on their own (252; 255). As such, the TMT may provide useful supplemental data for this dissertation project.

Freestanding Response Validity Tests

Victoria Symptom Validity Test (VSVT)

The VSVT is a computerized, freestanding response validity test that uses a forced-choice, digit recognition paradigm to assess the possible feigning or exaggeration of cognitive impairments (243). On forty-eight trials (three blocks of sixteen trials), participants are shown a five-digit sequence, and then are asked to choose between two options: 1) the correct five-digit number, or 2) a foil (i.e., a similar but different digit sequence). The items are categorized as either "easy" or "difficult," depending on the similarity of the foil to the correct five-digit number. The test measures the Total Items Correct score, which includes the type and number of items answered correctly, Response Latency, and Right-Left Preference scores. Like other forced-choice measures, the VSVT is interpreted based on the comparison between the actual score and what is

expected to occur on chance alone. The assessment takes approximately 18 to 25 minutes to administer (243; 261).

In a recent meta-analysis of freestanding RVTs, Sollman and Berry (248) reported the combined effect size of the VSVT Hard index in differentiating valid and invalid responding groups was *d*=2.77 (95% *CI*: 2.32-3.22), significantly higher than the Word Memory Test (WMT; 93), Test of Memory Malingering (TOMM; 263), the Letter Memory Test (LMT; 130), and the Medical Symptom Validity Test (MSVT; 94). Furthermore, Sollman and Berry (248) also reported a VSVT Difficult Items Correct cutoff score of 15 or below produced an average specificity of 95.5% (95% *CI*: 76.4-100%) and average sensitivity of 81.5% (95% *CI*: 75.1-87.9%). Additionally, VSVT failure rates have been consistently shown to be higher in compensation-seeking samples than clinical populations (103). Available data suggest the VSVT is relatively unaffected by psychosis, as well as depression of all severities and subtypes (90). The VSVT appears to be a highly sensitive and specific freestanding RVT.

Medical Symptom Validity Test (MSVT)

The MSVT is a computerized, forced-choice response validity test used to determine cognitive effort and the possible feigning or exaggeration of symptoms (94). During the MSVT, a list of word pairs (i.e., “skipping” and “rope”) is presented twice on the computer screen. The participant is then asked to choose the correct word from a pair of words, one being the target and one being the foil. After a 10-minute delay, the participant performs the forced-choice task again, and then completes the paired associates and free recall subtests (94; 126).

In Sollman and Berry's (248) meta-analysis of freestanding RVTs, the MSVT pass/fail index had a combined effect size (*d*) of 0.94 (95% *CI*: 0.70-1.19) when differentiating valid and invalid responding groups[14]. The authors (248) also reported that the MSVT pass/fail index demonstrated good-to-excellent classification accuracy for invalid responding, producing an average specificity of 91.3% (95% *CI*: 64.1-100%) and average sensitivity of 70.0% (95% *CI*: 13.1-100.0%). The MSVT's comparative ease-of-use and classification accuracy make it a useful supplemental freestanding RVT for this project.

PROCEDURE

As described earlier, this dissertation project incorporated a prospective simulator study and data collected from this dissertation project's parent study. The simulator study included two groups of persons without a history of traumatic brain injury: a group biased to perform as if they sustained a head injury (biased responding; BR) and a control group asked to perform their best (unbiased responding; UR). The parent study's data provided a clinical comparison group of persons with self-reported history of mild TBI who did not meet criteria for invalid responding (unbiased responding with mild TBI; UR-mTBI). Research staff was divided into two groups, study coordinators and examiners. Study coordinators were responsible for assigning participants to groups, collecting pre-test data, and conducting post-test interviews. Examiners were responsible for administering and scoring the neurocognitive battery, and they were blinded to group assignment. The following section will describe the procedure for the prospective, simulator study, and will also describe the procedure used in the parent study.

[14] Surprisingly, this large effect size was significantly lower than the effect sizes found for the VSVT, WMT, TOMM, and LMT.

Simulator Study: BR and UR Groups

Potential participants were screened via a phone interview in which head injury history, neurological illness history, and demographic information were obtained by a lab member. If the participant met inclusion criteria and did not meet exclusion criteria, the participant was scheduled to come to the lab and complete the simulator study neurocognitive assessment battery. After the phone interview—but before the participant arrived for assessment—a study coordinator assigned the participant a study identification number. Study identification numbers began at "1" and increased sequentially as additional participants were scheduled. Using a random group assignment plan created from a randomly permuted block assignment generator program (60), the study coordinator then assigned the participant to either the biased responding (BR) or unbiased responding (UR) group. Participants were not told of their group assignment in advance of assessment date in order to mitigate potentially confounding effects of test preparation or coaching (241).

Once the participant arrived for testing, a study coordinator obtained their informed consent. If the participant agreed to participate, a study coordinator administered the baseline interview, WTAR, and HIKS Version B[15]. A study coordinator then presented a specific group assignment script (see Appendix C: Simulator Study Group Assignment Scripts) to the participant and asked the participant not to reveal group membership to the examiner. The group assignment script was adapted from previous studies of invalid responding (71; 251; 257; 279), and was designed to ask both BR and UR group members to act as if they were involved in a remote vehicular accident. While both scripts stated that the participants do not feel any lingering effects

[15] HIKS Version B is designed for persons who have not sustained a TBI, and was given to all simulator study participants regardless of group assignment.

from the accident, the biased responding script asked participants to exaggerate cognitive problems in order to get money from an insurance company. Furthermore, the biased responding scenario warned participants to fake believably[16]. By contrast, the unbiased responding scenario asked participants to perform their best. After the group scenarios were presented to the participants, a study coordinator addressed the participant's questions, comments, or concerns, if necessary.

When ready, the participant was introduced to his or her examiner and completed a 1.5 hour neurocognitive assessment battery. Examiners consisted of clinical psychology doctoral students who have completed graduate courses and lab-internal training on clinical assessment. All examiners' testing competence and protocol adherence were verified by Dr. Ettenhofer and this author prior to assessing any study participants. The assessment proceeded in the following order: 1) BEAM, 2) VSVT, 3) Digit Span, 4) CPT-II, 5) MSVT, 6) TMT A & B, 7) King-Devick Test[17], and 8) Neurobehavioral Symptom Inventory[18]. Once the neurocognitive assessment was completed, the examiner thanked the participant for taking part in the study and left the room. A study coordinator who knew the participant's assigned group would then administer a group-assignment-specific feedback interview (see Appendix C: Simulator Study Feedback Interview) and debrief script (see Appendix C: Simulator Study Debrief Scripts) to the participant.

[16] In their meta-analysis of 38 studies of invalid responding, Sollman and Berry (2011) reported that warnings to fake believably significantly increased the effect size of freestanding RVT score differences between valid and invalid responding groups of healthy simulators.

[17] The King-Devick Test was added for secondary analyses not included in this project's aims.

[18] The Neurobehavioral Symptom Inventory was added for secondary analyses not included in this project's aims.

Parent Study: UR-mTBI Group

The UR-mTBI group data were collected from this dissertation project's parent study. Participants who met criteria for inclusion in the TBI cohort were scheduled for testing with a lab examiner. Examiners consisted of clinical psychology doctoral students who completed graduate courses and lab-internal training on clinical assessment. All examiners' testing competence and protocol adherence were verified by Dr. Ettenhofer prior to assessing any study participants. Participants that were eligible for $40 compensation were made aware that the compensation was fixed prior to their arrival for testing.

On the day of testing, the examiners administered a semi-structured interview that asked, among other things, about the participant's history of brain injury, to include most recent/most severe injury information, mechanism of injury, loss of consciousness length, and posttraumatic amnesia length, among others. Current level of fatigue, alcohol/nicotine/caffeine consumption within previous 12 hours, and medication information was also collected. After the semi-structured interview, the UR-mTBI group participants were administered the BEAM and a comprehensive neurocognitive battery that measured domains of attention, executive function, memory, processing speed, and psychomotor ability. The participants were asked to perform their best throughout the battery, and all attempts were made by examiners to eliminate sources of response bias.

Since the archival UR-mTBI group data were drawn from a parent study with an established assessment battery, not all of the measures from the prospective simulator study could be compared to the UR-mTBI group. Specifically, there were no freestanding response validity tests to compare between BR, UR, and UR-mTBI groups. However, there were several measures with empirically validated embedded response

validity indices that were used to identify "valid" and "invalid" responders within the parent study's TBI cohort. Of the neurocognitive assessments given to the BR and UR groups, the UR-mTBI groups had group comparison data for the following: 1) Baseline interview, 2) BEAM, 3) WTAR, 4) Digit Span, 5) CPT-II, 6) TMT, and 7) NSI[19].

Data Analysis

General Analytic Strategy

Data from the simulator study were analyzed to address Specific Aims 1 and 2. Data from the parent study were compared with simulator study data to address Specific Aim 3. First, descriptive statistics were calculated for the two simulator study groups: biased responders without a history of brain injury (BR) and unbiased responders without a history of brain injury (UR). Demographics (e.g., age, sex, race/ethnicity, years of education, etc.), estimated premorbid intelligence (i.e., WTAR), and knowledge of head injuries (i.e., HIKS) were compared using chi square analyses or *t*-tests to identify any significant demographic differences between the BR and UR groups.

To address Specific Aims 1 and 2, receiver operating characteristics (ROC) analyses and logistic regressions were performed to identify BEAM, embedded RVT, and freestanding RVT variables with the greatest potential to differentiate between groups of biased and unbiased responders (64; 110; 127; 168; 172). Variables with area under the curve (AUC) greater than or equal to 0.7 (acceptable classification accuracy) and *p* values less than or equal to .05 were identified. Due to the large number of variables meeting this criteria, only variables with AUC greater than or equal to 0.9 (i.e., outstanding classification accuracy) were submitted to subsequent analyses. Shapiro-Wilk and

[19]NSI data were considered secondary to this dissertation project's aims and were collected for future studies.

Levene's tests were performed to test for normality and homogeneity of variance, respectively. Independent-samples *t*-tests and Mann-Whitney *U* tests were then conducted to identify group differences and effect sizes. Classification accuracy statistics—sensitivity, specificity, hit rate, positive predictive value, negative predictive value, and likelihood ratios—were calculated for cutoff scores that approximated 90% specificity or higher. Stepwise logistic regression analyses were conducted on variables with outstanding classification accuracy in order to determine the best and most representative variables among the BEAM, embedded RVTs, and freestanding RVTs. Hierarchical logistic regressions were conducted on these representative variables to determine incremental predictive value of the BEAM above and beyond embedded and freestanding RVTs.

To address Specific Aim 3, the variables that demonstrated outstanding classification accuracy were compared to archival data of unbiased responders with a history of mild traumatic brain injury (UR-mTBI). Shapiro-Wilk and Levene's tests were performed to test for normality and homogeneity of variance, respectively. One-way ANOVA and Kruskal-Wallis tests were performed to identify omnibus group differences among the BR, UR, and UR-mTBI groups. Post-hoc Tukey HSD and Mann-Whitney *U* tests were performed to identify group differences and effect sizes between the three groups. Sensitivity (SN) and specificity (SP) were calculated for UR-mTBI cutoff scores that approximated 90% specificity or higher.

Control Variables

Age, gender, and years of education often influence normative values in neurocognitive assessment (114). Variables obtained from the WTAR, TMT, CPT-II,

and Digit Span can be corrected for one or more of these demographic variables using normative data (55; 114; 278). However, BEAM variables are not corrected for demographic variables, and cannot be meaningfully compared to demographically-adjusted scores from other neurocognitive tests. As such, uncorrected or "raw" values were used for all primary analyses.

Data Analytic Strategy for Aims and Hypotheses

Specific Aim 1

To support Specific Aim 1 of this dissertation project, BEAM data were analyzed to determine which variables demonstrated the best ability to differentiate valid from invalid responding. Since the BEAM is a novel test of cognitive functioning, little was currently known about its ability to detect invalid responding. As such, twenty-nine BEAM variables (e.g., saccadic, manual, and data validity metrics) were submitted to ROC analyses. Twenty-five BEAM variables met the initial criteria of AUC ≥ 0.7 and $p < .05$. To reduce the risk of Type I error when identifying statistically significant variables (167; 202), only variables with AUC greater than or equal to 0.9 and p values less than 0.05 were selected for additional analyses. It should be noted that AUC estimates from ROC curves are similar for a wide range of normal and non-normal distributions (109), so the default parametric ROC analyses available in SPSS version 20 were conducted. Classification accuracy tables were prepared. Cutoff scores that approximated 90% specificity or higher were calculated for all BEAM variables that demonstrated outstanding classification accuracy.

Based on results of several studies of feigned neurocognitive performance (42; 129; 138; 188; 282), the likelihood of obtaining non-normal kurtosis and negative skew

among the BEAM variables was high. Additionally, each group in the study had a sample size less than 50. As such, all BEAM variables were assessed for normality using Shapiro-Wilk tests (208; 238). For variables with non-significant Shapiro-Wilk test results, independent sample *t*-tests were used to test the null hypothesis that there were no significant differences between BR and UR groups. Between-group effect sizes (Cohen's *d*) were determined for all comparisons between normally distributed variables. For variables whose Shapiro-Wilk tests revealed significantly non-normal data distributions, non-parametric Mann-Whitney *U* tests were conducted to test the null hypothesis that there were significant differences between BR and UR groups. Between-group effect sizes were calculated for non-normally distributed variables using the *r* statistic (i.e., the Z score obtained from Mann-Whitney *U* test divided by the square root of the total sample size).

Specific Aim 2

To support Specific Aim 2 of this dissertation project, embedded and freestanding RVT variables were analyzed in manner described above. Of the eighteen embedded RVT variables submitted to ROC analyses, thirteen met the initial criteria of AUC ≥ 0.7 and $p < .05$. All fifteen freestanding RVT variables submitted to ROC analyses met the initial criteria of AUC ≥ 0.7 and $p < .05$. Accordingly, only embedded and freestanding RVT variables that demonstrated AUC ≥ 0.9 were submitted to tests of normality, homogeneity of variance, and group differences. Effect sizes were calculated for all embedded and freestanding RVT variables with significant differences between BR and UR groups.

Next, all BEAM, embedded RVT, and freestanding RVT variables with AUC ≥ 0.9 were submitted to a series of stepwise logistic regressions in order to determine the variable from each test type that best predicted simulator study group membership. Forward and backward stepwise logistic regressions were used to efficiently reduce the number of variables for each test type. If a model's chi square value was significant at the $p < .05$ level, variables with significant ($p < .05$) Wald chi square statistics were retained for subsequent stepwise logistic regressions. Once a "representative" variable was identified for each test type (i.e., BEAM, embedded RVT, freestanding RVT), a series of hierarchical logistic regression models were performed to assess the BEAM's incremental predictive value above and beyond existing response validity measures. Unlike stepwise logistic regression models, which treat variables equally, hierarchical logistic regressions assume the variables are loaded in order of predictive value. Accordingly, three analyses were performed: 1) BEAM above and beyond Embedded RVTs; 2) BEAM above and beyond Freestanding RVTs; and 3) BEAM above and beyond both Embedded and Freestanding RVTs.

For the first hierarchical logistic regression model, the representative Embedded RVT variable was loaded into block 1 ("Embedded") and compared between BR and UR groups. Chi square values, correct classification percentages, and exponent *B* values were calculated. In the next step, the representative BEAM variable was loaded into block 2 ("BEAM"). As before, chi square values, correct classification percentages, and exponent *B* values were calculated. Differences in chi square values between block 1 (i.e., Embedded) and block 2 (i.e., BEAM) were also be calculated to determine incremental predictive value of BEAM metrics above and beyond the classification

accuracy of embedded RVTs used in this study. A second hierarchical logistic regression model comparing the representative Freestanding RVT variable in block 1 to the representative BEAM variable in block 2 was also performed.

A third and final model was used to determine if the representative BEAM variable provided incremental predictability of invalid responding above and beyond the representative variables from both embedded and freestanding RVT. The order of blocks was as follows: 1) Embedded RVT variable; 2) Freestanding RVT variable; and 3) BEAM variable. Changes in chi square values on step 2 indicated whether or not the representative freestanding RVT variable provided incremental predictive value of invalid responding above and beyond the representative embedded RVT variable. Changes in chi square values on step 3 indicated whether or not the representative BEAM variable provided significant predictive value above and beyond the representative variables from both embedded and freestanding RVTs.

Specific Aim 3

To support Specific Aim 3 of this dissertation project, BEAM and embedded RVT variables that demonstrated outstanding classification accuracy in the simulator study were compared to archival data of unbiased responders with a history of mild traumatic brain injury (UR-mTBI). First, the mild TBI cohort from this project's parent study was screened for meeting exclusion criteria. Next, chi-square analyses and ANOVAs were used to compare the UR and BR groups to the UR-mTBI group on factors of age, gender, years of education, estimated premorbid intelligence, and knowledge of head injuries to identify any variables that needed to be controlled during subsequent analyses.

Shapiro-Wilk and Levene's tests were performed to identify variables that violated assumptions of normal distributions and homogeneity of variance, respectively, in the UR-mTBI group. One-way ANOVA and Kruskal-Wallis tests were then performed to identify omnibus group differences among the BR, UR, and UR-mTBI groups. Post-hoc Tukey HSD and Mann-Whitney *U* tests were performed to identify group differences and effect sizes between the three groups. Sensitivity to invalid responding (i.e., the BR group) was calculated for UR-mTBI cutoff scores that demonstrated 85% specificity or higher. Given the absence of an "invalid responding" clinical comparison group, predictive value statistics (e.g., PPV, NPV) were not calculated in this step.

Alpha Level

A two-tailed alpha level of .05 was used for all ROC analyses. To mitigate the likelihood of making a Type 1 error in subsequent between-groups analyses, only variables with AUC ≥ 0.7 were expected to be considered for subsequent analyses. However, the majority of the BEAM, embedded RVT, and freestanding RVT variables met this criteria, and a new cutoff of AUC ≥ 0.9 was implemented to reduce Type I error. Due to the exploratory nature of the study, correction factors were not planned at this time of this dissertation project's proposal. However, the relatively high number of variables that met the AUC ≥ 0.9standard prompted this author to utilize Bonferroni corrections for multiple groups. As such, all between-groups analyses for Specific Aims 1 and 2 (BR and UR groups) used a Bonferroni-corrected .002 level of significance, and between-groups analyses for Specific Aim 3 (BR, UR, and UR-mTBI groups) used a Bonferroni-corrected .003 level of significance.

Power Analysis

Multiple meta-analyses of embedded and freestanding response validity tests have found Cohen's *d* effect sizes greater than 1.0—a very large group difference—between two groups of invalid and valid responders (137; 248; 272). Armistead-Jehle and Buican (9) reported large effect sizes across neurocognitive abilities as a function of RVT performance, especially in tests of attention, processing speed, and memory. These large effect sizes influence the sample size required to sufficiently power this dissertation project's analyses.

Power analyses were computed using G*Power Version 3 (75). At the time of this project's proposal, it was estimated that at least 20 subjects per group (BR, UR, UR-mTBI) would be required to obtain 80% power. Data collection for this project was conducted from June 2013 until December 2013. All efforts were made to meet and exceed the planned sample sizes for the three groups. The simulator study's final sample size of 50 (n = 24 in BR group; n = 26 in UR group) was found to sufficiently power parametric (i.e., independent samples *t*-tests) and non-parametric (i.e., Mann-Whitney *U* tests) between-groups analyses in Specific Aims 1 and 2 (75). The final UR-mTBI sample size (n = 19), despite being one subject below the planned sample size, was found to sufficiently to power parametric (i.e., one-way ANOVAs) and non-parametric (i.e., Kruskal-Wallis tests) between-groups analyses in Specific Aim 3 (75).

CHAPTER 3: Results

Fifty-seven people contacted study coordinators to participate in the prospective, experimental study of the dissertation project. Of these, one was ineligible due to history of concussion and five could not participate due to scheduling conflicts. Of the fifty-one participants who attempted to complete the experiment, one did not complete the protocol due to technical difficulties with the lab equipment. As a result, fifty participants completed the experiment.

All study participants were randomly assigned to either the Biased Responder (BR) group or the Unbiased Responder (UR) group. Demographic information for these groups is shown in Table 3. The BR group (n = 24) and UR group (n = 26) did not differ significantly on age, years of formal education, estimated premorbid intelligence, gender, or race/ethnicity. Additionally, the BR group did not significantly differ from the UR group on knowledge of head injury sequelae prior to group assignment.

SPECIFIC AIM 1

Examining relationships between invalid responding and performance on BEAM metrics (Hypotheses 1A-1E)

ROC analyses were conducted to identify the accuracy of each BEAM variable in predicting membership in the BR or UR groups. Twenty-nine BEAM variables were submitted to ROC analyses: one from the number of trials with invalid initial fixations (Hypothesis 1A), twelve from saccadic reaction time (SacRT) and manual reaction time (ManRT) for the overall measure and five trial types (Hypothesis 1B), twelve from saccadic and manual RT intra-individual variability (IIV) for the overall measure and five

trial types[20] (Hypothesis 1C), two from saccadic and manual commission error percentage (Com%) from the inhibition trial type[21] (Hypothesis 1D), and two from the saccadic and manual omission error percentage (Om%) from the non-inhibition trial types (Hypothesis 1E).

As stated earlier in the manuscript, at least 10 successfully recorded trials per trial type were required to generate BEAM summary metrics. Several participants in the biased responding group performed in such a manner where the BEAM could not calculate summary reaction time metrics (e.g., too many omissions, too many trials with data loss). As a result, a small amount of BEAM data (1.9% of all BEAM data) was missing, exclusively from the BR group. The results from the ROC analyses are shown in Table 4.

Twenty-six BEAM variables demonstrated statistically significant classification accuracy ($p < .05$); twenty-five BEAM variables demonstrated acceptable classification accuracy (area under the curve [AUC] ≥ 0.70). The Number of Invalid Fixations variable was statistically significant ($p = .02$) but demonstrated less-than-acceptable classification accuracy (AUC = 0.69; 95% *CI* = 0.53 to 0.84). As such, Hypothesis 1A was *not confirmed*. SacRT-UCG ($p = .19$; AUC = 0.61; 95% *CI* = 0.44 to 0.78), SacRT-UC ($p = .25$; AUC = 0.60; 95% *CI* = 0.42 to 0.77), and SacOm% ($p = .25$; AUC = 0.60; 95% *CI* = 0.44 to 0.75) failed to reach statistical significance and demonstrated classification accuracy below the acceptable range.

[20] The five trial types with RT and RT-IIV (i.e., the non-inhibition trial types) are Directional Cue (DC), Nondirectional Cue (NDC), Misdirectional Cue (MDC), Uncued with Gap (UCG), and Uncued (UC).

[21] The inhibition trial type is Directional Cue-Red Arrow (DCR)

ROC analyses indicated that ten of the twelve BEAM RT variables achieved acceptable-to-excellent classification accuracy (AUC range: 0.71-0.89). As such, Hypothesis 1B was *partially confirmed*. Hypothesis 1C was *confirmed*, as all twelve BEAM RT-IIV variables demonstrated excellent-to-outstanding classification accuracy (AUC range: 0.85-0.97). Hypothesis 1D was *confirmed*, as both SacCom% (AUC = 0.94; 95% *CI* = 0.87 to 1.00) and ManCom% (AUC = 0.80; 95% *CI* = 0.66 to 0.93) variables demonstrated excellent-to-outstanding classification accuracy. Lastly, Hypothesis 1E was *partially confirmed*, as ManOm% (AUC = 0.94; 95% *CI* = 0.87 to 1.00) demonstrated outstanding classification accuracy while SacOm% demonstrated unacceptable classification accuracy.

Of the ten BEAM RT variables with AUC ≥ 0.7, four were saccadic RT and six were manual RT. SacRT-MDC (AUC = 0.81; 95% *CI* = 0.69 to 0.94) demonstrated excellent classification accuracy, while SacRT-DC (AUC = 0.77; 95% *CI* = 0.65 to 0.90), SacRT-Overall (AUC = 0.74; 95% *CI* = 0.60 to 0.88), and SacRT-NDC (AUC = 0.71; 95% *CI* = 0.57 to 0.86) demonstrated acceptable classification accuracy. All manual RT variables demonstrated excellent classification accuracy: ManRT-Overall (AUC = 0.89; 95% *CI* = 0.80 to 0.99), ManRT-UCG (AUC = 0.89; 95% *CI* = 0.78 to 0.99), ManRT-NDC (AUC = 0.88; 95% *CI* = 0.78 to 0.98), ManRT-MDC (AUC = 0.86; 95% *CI* = 0.74 to 0.99), ManRT-DC (AUC = 0.84; 95% *CI* = 0.73 to 0.96), and ManRT-UC (AUC = 0.83; 95% *CI* = 0.70 to 0.95).

All twelve (six saccadic and six manual) BEAM RT-IIV variables achieved excellent-to-outstanding levels of classification accuracy. SacRT-IIV-Overall (AUC = 0.97; 95% *CI* = 0.93 to 1.00), ManRT-IIV-Overall (AUC = 0.97; 95% *CI* = 0.92 to 1.00),

ManRT-IIV-DC (AUC = 0.97; 95% *CI* = 0.92 to 1.00), ManRT-IIV-NDC (AUC = 0.96; 95% *CI* = 0.90 to 1.00), ManRT-IIV-MDC (AUC = 0.96; 95% *CI* = 0.90 to 1.00), SacRT-IIV-DC (AUC = 0.93; 95% *CI* = 0.86 to 1.00), ManRT-IIV-UCG (AUC = 0.93; 95% *CI* = 0.84 to 1.00), SacRT-IIV-MDC (AUC = 0.92; 95% *CI* = 0.83 to 1.00), ManRT-IIV-UC (AUC = 0.92; 95% *CI* = 0.84 to 0.99), and SacRT-IIV-UCG (AUC = 0.90; 95% *CI* = 0.80 to 1.00) demonstrated outstanding classification accuracy. SacRT-IIV-UC (AUC = 0.85; 95% *CI* = 0.75 to 0.96), and SacRT-IIV-NDC (AUC = 0.85; 95% *CI* = 0.74 to 0.96) demonstrated excellent classification accuracy.

To reduce the number of variables submitted to further analyses, only variables that demonstrated outstanding classification accuracy (AUC $\geq$ 0.9) were considered. Using this criterion, twelve BEAM variables were retained: SacRT-IIV-Overall, ManRT-IIV-Overall, ManRT-IIV-DC, ManRT-IIV-NDC, ManRT-IIV-MDC, ManOm%, SacCom%, SacRT-IIV-DC, ManRT-IIV-UCG, SacRT-IIV-MDC, ManRT-IIV-UC, and SacRT-IIV-UCG. Notably, ten of the twelve outstanding BEAM variables were reaction time intra-individual variability. The other two BEAM variables with outstanding classification accuracy were SacCom% and ManOm%.

Shapiro-Wilk tests of normality revealed significantly non-normal distributions for four BEAM variables: ManRT-IIV-UCG (BR group: $W(21) = .87$, $p = .01$), ManRT-IIV-UC (BR group: $W(23) = .90$, $p = .03$), SacCom% (UR group: $W(26) = .81$, $p < .001$), and ManOm% (UR group: $W(26) = .28$, $p < .001$; BR group: $W(24) = .77$, $p < .001$). As shown in Table 5, Mann-Whitney *U* tests with a Bonferroni correction of .002 level of significance indicated biased responders performed significantly worse (i.e., higher error rate, greater RT variability) than unbiased responders on the following variables:

ManRT-IIV-UCG (U = 41.0, p < .001, r = .72), ManRT-IIV-UC (U = 51.0, p < .001, r = .71), ManOm% (U = 36.5, p < .001, r = .78), and SacCom% (U = 39.0, p < .001, r = .75). Very large effect sizes were found for all between-group comparisons (r range: .71 - .78).

Shapiro-Wilk tests for the other eight BEAM variables with outstanding classification accuracy were not significant (p > .05) in either BR or UR group. Independent samples t-tests were conducted on the eight BEAM variables with normal distributions. A Bonferroni correction was applied so all effects are reported at a .002 level of significance. As shown in Table 6, the t-tests indicated that reaction time intra-individual variability was significantly greater in the BR group than the UR group for the following variables: SacRT-IIV-DC, t(48) = 7.72; SacRT-IIV-Overall, t(48) = 9.49; ManRT-IIV-DC, t(47) = 10.2; ManRT-IIV-NDC, t(47) = 8.99; ManRT-IIV-MDC, t(42) = 7.49; ManRT-IIV-Overall, t(47) = 10.2; SacRT-IIV-MDC, t(32) = 7.21; and SacRT-IIV-UCG, t(33) = 6.86. Levene's test indicated unequal variances for SacRT-IIV-MDC (F = 7.64, p = .008) and SacRT-IIV-UCG (F = 4.19, p = .046), so degrees of freedom were adjusted from 48 to 32 and 47 to 33, respectively. Very large effect sizes were found for all between-group comparisons (Cohen's d range: 1.99 - 2.90). Classification accuracy statistics for the twelve best BEAM variables are shown in Table 7.

A series of stepwise logistic regressions were conducted to identify the best predictors of invalid responding among the twelve BEAM variables. To keep from overfitting the regression models, the twelve BEAM variables were divided into five groups of similar variables:

1) SacCom% and ManOm%;

2) SacRT-IIV-DC, SacRT-IIV-MDC, and SacRT-IIV-UCG,

3) ManRT-IIV-DC, ManRT-IIV-NDC, and ManRT-IIV-MDC;

4) ManRT-IIV-UCG and ManRT-IIV-UC; and

5) SacRT-IIV-Overall and ManRT-IIV-Overall.

All intra-individual variability variables were converted to milliseconds in order to obtain interpretable odds ratios. Significant predictor variables from each group were identified and loaded into subsequent stepwise regressions until the best predictors of the biased responder group were identified (see Figure 1).

Analyses indicated that SacRT-IIV-Overall and ManRT-IIV-Overall as a set most reliably differentiated the BR and UR groups, $\chi^2(2, N = 49) = 53.6, p < .001$. Nagelkerke's R^2 of .89 indicated a strong relationship between prediction and grouping. Prediction success overall was 95.9% (96.2% for UR group and 95.7% for BR group). Both SacRT-IIV-Overall (χ^2 (1) = 4.80, p = .03) and ManRT-IIV-Overall (χ^2 (1) = 5.36, p = .02) significantly contributed to group prediction. Exp(B) values indicated that when SacRT-IIV-Overall and ManRT-IIV-Overall increases by one millisecond, a person is 1.09 and 1.08 times more likely to be an invalid responder. Put a different way, every 10 millisecond increase in the standard deviation of an individual's average overall saccadic or manual reaction time increases the chances of a person being an invalid responder by more than 2 times. Additional ROC analyses indicated that the combined AUC of SacRT-IIV-Overall + ManRT-IIV-Overall was nearly perfect (AUC = 0.99; 95% *CI* = 0.96 to 1.00). SacRT-IIV-Overall and ManRT-IIV-Overall each contributed an additional 0.02 AUC to the model above and beyond their individual classification accuracy.

SPECIFIC AIM 2

Comparing the invalid responding classification accuracy abilities of BEAM metrics to existing response validity tests (Hypotheses 2A-2C).

Embedded RVTs

Eighteen variables from embedded response validity tests (RVTs) were submitted to ROC analyses. As seen in Table 8, thirteen of these variables demonstrated acceptable-or-greater classification accuracy. TMT A time (AUC = 0.74; 95% *CI* = 0.59 to 0.90, *p* = .003) and B time (AUC = 0.79; 95% *CI* = 0.66 to 0.91, *p* = .001) demonstrated acceptable classification accuracy. Seven CPT-II variables demonstrated significant ($p < .001$) classification accuracy that ranged from acceptable to outstanding: RT ISI Change (AUC = 0.79; 95% *CI* = 0.67 - 0.91), Perseverations (AUC = 0.84; 95% *CI* = 0.73 to 0.96), Detectability (AUC = 0.88; 95% *CI* = 0.77 to 0.98), Variability (AUC = 0.89; 95% *CI* = 0.80 to 0.98), Hit RT Standard Error (AUC = 0.90; 95% *CI* = 0.82 to 0.99), Omissions (AUC = 0.91; 95% *CI* = 0.82 to 1.00), and Commissions (AUC = 0.93; 95% *CI* = 0.85 to 1.00). All four variables from the WAIS-IV Digit Span subtest, Age-Corrected Scaled Score (ACSS; AUC = 0.94; 95% *CI* = 0.88 to 1.00), Reliable Digit Span-Revised (RDS-R; AUC = 0.94; 95% *CI* = 0.88 to 1.00), Alternative Reliable Digit Span (ARDS; AUC = 0.94; 95% *CI* = 0.86 to 1.00), and Reliable Digit Span (RDS; AUC = 0.93; 95% *CI* = 0.84 to 1.00), demonstrated outstanding and significant ($p < .001$) classification accuracy. CPT-II Standard Error Interstimulus Interval Change (AUC = 0.67; 95% *CI* = 0.52 to 0.82, *p* = .04) was significant but demonstrated below-acceptable classification accuracy. CPT-II Hit RT (AUC = 0.52; 95% *CI* = 0.35 to 0.69, *p* = .83), Hit RT Block Change (AUC = 0.50; 95% *CI* = 0.33 to 0.66, *p* = .96), Hit RT Standard Error Block Change (AUC = 0.54; 95% *CI* = 0.37 to 0.70, *p* = .67), and Response Style (AUC = 0.67; 95% *CI*

= 0.51 to 0.82, p = .05) did not significantly differentiate between biased and unbiased responders.

As with the BEAM variables in Aim 1, only embedded RVT variables with outstanding (AUC ≥ 0.9) classification accuracy were submitted to further analyses. Of these seven variables, four were from the WAIS-IV Digit Span: Reliable Digit Span-Revised (RDS-R), Alternative Reliable Digit Span (ARDS), Age-Corrected Scaled Score (ACSS), and Reliable Digit Span (RDS). Three variables from the CPT-II—Omissions, Commissions, and Hit RT Standard Error (SE)—were also included in subsequent analyses.

Shapiro-Wilk tests indicated that RDS-R (UR group: $W(26) = .90$, $p = .015$), ARDS (UR group: $W(26) = .92$, $p = .040$), CPT-II Omissions (UR group: $W(26) = .64$, $p < .001$; BR group: $W(24) = .66$, $p < .001$), and CPT-II Hit RT SE (BR group: $W(24) = .80$, $p < .001$) were significantly non-normal. As shown in Table 9, Mann-Whitney U tests with a Bonferroni correction of .002 level of significance indicated that RDS-R scores in the BR group (Mdn = 11.0) were significantly lower than the UR group (Mdn = 17.0), U = 35.0, $p < .001$, r = .77, and that ARDS scores in the BR group (Mdn = 8.50) were significantly lower than the UR group (Mdn = 12.0), U = 38.0, $p < .001$, r = .76. Additional Mann-Whitney U tests indicated that the BR group (Mdn = 6.50) made significantly more Omission errors on the CPT-II than the UR group (Mdn = 0.00), U = 57.5, $p < .001$, r = .72, and that Hit RT SE was significantly higher in the BR group (Mdn = 7.56) than the UR group (Mdn = 4.20), U = 61.0, $p < .001$, r = .69.

Shapiro-Wilk tests for ACSS, RDS, and CPT-II Commissions were not significant ($p > .05$) in either the BR or UR group. Independent samples t-tests were conducted on

the embedded RVT variables with normal distributions. A Bonferroni correction was applied so all effects are reported at a .002 level of significance. As shown in Table 10, independent samples t-tests on WAIS-IV Digit Span variables indicated that the BR group had significantly lower ACSS, $t(48) = 7.83$, $p < .001$, and RDS, $t(48) = 6.91$, $p < .001$, scores than the UR group. Additionally, the BR group made significantly more Commission errors, $t(48) = 7.72$, $p < .001$, on the CPT-II than the UR group. Very large effect sizes were found for all between-group comparisons among the normally distributed embedded RVT variables (Cohen's *d* range: 1.94-2.23). Classification accuracy statistics for the seven best embedded RVT variables are shown in Table 11.

A series of stepwise logistic regressions were conducted to identify the best predictors of invalid responding among the seven embedded RVT variables. To keep from overfitting the models, the seven embedded RVT variables were broken into two groups of similar variables, one from the WAIS-IV Digit Span subtest (ACSS, RDS, RDS-R, and ARDS), and one from the CPT-II (Omissions, Commissions, and Hit RT SE). Significant predictor variables from each group were identified and loaded into subsequent stepwise regressions until the best predictors of the biased responder group were identified.

Analyses indicated that RDS-R and CPT-II Commissions as a set most reliably differentiated the BR and UR groups, χ^2 (2, $N = 50$) = 49.6, $p < .001$. Nagelkerke's R^2 of .84 indicated a strong relationship between prediction and grouping. Prediction success overall was 94.0% (96.2% for UR group and 91.7% for BR group). CPT-II Commissions (χ^2 (1) = 5.95, $p = .02$) and RDS-R (χ^2 (1) = 7.11, $p = .008$) both made significant contributions to group prediction. Exp(B) values indicated that each additional CPT-II

Commission error increases the likelihood of a person being an invalid responder by 1.33 times. Since higher RDS-R scores are consistent with better performance, each fewer RDS-R point increases the likelihood of a person being an invalid responder by 2.27 times. Additional ROC analyses indicated that the combined AUC of RDS-R + CPT-II Commissions was 0.97 (95% *CI* = 0.93 to 1.00). RDS-R contributed an additional 0.046 to the overall classification accuracy above and beyond CPT-II Commissions, and CPT-II Commissions contributed an additional 0.027 to the classification accuracy above and beyond RDS-R.

Freestanding RVTs

Fifteen variables from freestanding RVTs were submitted to ROC analyses. As seen in Table 12, all fifteen variables demonstrated significant ($p < .001$) classification accuracy that ranged from excellent-to-outstanding. The VSVT's Total RT[22] SD (AUC = 0.87; 95% *CI* = 0.77 to 0.98), Easy RT SD (AUC = 0.85; 95% *CI* = 0.74 to 0.96), and Easy Correct (AUC = 0.85; 95% *CI* = 0.73 to 0.96) variables demonstrated excellent classification accuracy. MSVT Immediate Recognition % Correct (IR; AUC = 0.90; 95% *CI* = 0.80 to 1.00), MSVT Delayed Recognition % Correct (DR; AUC = 0.92; 95% *CI* = 0.84 to 1.00), MSVT Consistency % Correct (CNS; AUC = 0.94; 95% *CI* = 0.87 to 1.00), MSVT Paired Associates % Correct (PA; AUC = 0.93; 95% *CI* = 0.85 to 1.00), MSVT Free Recall % Correct (FR; AUC = 0.96; 95% *CI* = 0.91 to 1.00), MSVT Fail Any Subtest (AUC = 0.90; 95% *CI* = 0.80 to 1.00), VSVT Easy RT (AUC = 0.90; 95% *CI* = 0.81 to 0.98), VSVT Difficult RT (AUC = 0.91; 95% *CI* = 0.82 to 0.99), VSVT Difficult RT SD (AUC = 0.90; 95% *CI* = 0.81 to 0.99), and VSVT Total RT (AUC = 0.93; 95% *CI*

[22] Response Latency will be abbreviated as "RT" for consistency with other variables in this study

= 0.86 to 1.00) demonstrated outstanding classification accuracy. It should be noted that MSVT Fail Any Subtest is a dichotemous variable, unlike the fourteen other freestanding RVT variables. Two VSVT variables, Difficult Correct and Total Correct, were perfect (AUC = 1.00; 95% *CI* = 1.00 to 1.00) in their ability to identify invalid responding in this study's experimental sample.

Six of nine VSVT variables (Difficult Correct, Total Correct, Total RT, Difficult RT, Difficult RT SD, and Easy RT) and all six MSVT variables (Fail Any Subtest, IR, DR, CNS, PA, and FR) demonstrated outstanding classification accuracy and were submitted to additional analyses. Shapiro-Wilk tests of normality revealed significantly non-normal distributions for all eleven continuous, freestanding RVT variables: VSVT Difficult Correct (UR group: $W(26) = .68$, $p < .001$), VSVT Total Correct (UR group: $W(26) = .71$, $p < .001$), VSVT Easy RT (UR group: $W(26) = .75$, $p < .001$; BR group: $W(24) = .87$, $p = .004$), VSVT Difficult RT (UR group: $W(26) = .92$, $p = .04$; BR group: $W(24) = .78$, $p < .001$), VSVT Total RT (UR group: $W(26) = .87$, $p = .004$; BR group: $W(24) = .82$, $p = .001$), VSVT Difficult RT SD (UR group: $W(26) = .82$, $p < .001$; BR group: $W(24) = .79$, $p < .001$), MSVT IR (BR group: $W(24) = .89$, $p = .01$), MSVT DR (UR group: $W(26) = .38$, $p < .001$), MSVT CNS (UR group: $W(26) = .38$, $p < .001$), MSVT PA (UR group: $W(26) = .20$, $p < .001$), and MSVT FR (UR group: $W(26) = .85$, $p = .002$).

As shown in Table 13, Mann-Whitney *U* tests with a Bonferroni correction of .002 level of significance indicated statistically significant differences between the UR and BR groups on VSVT Difficult Correct ($U = 0.00$, $p < .001$, $r = .87$), VSVT Total Correct ($U = 0.00$, $p < .001$, $r = .87$), VSVT Easy RT ($U = 63.0$, $p < .001$, $r = .68$), VSVT

Difficult RT (U = 59.0, p < .001, r = .69), VSVT Difficult RT SD (U = 64.0, p < .001, r = .68), VSVT Total RT (U = 44.5, p < .001, r = .73), MSVT IR (U = 65.0, p < .001, r = .78), MSVT DR (U = 48.0, p < .001, r = .78), MSVT CNS (U = 37.5, p < .001, r = .81), MSVT PA (U = 42.0, p < .001, r = .82), and MSVT FR (U = 28.0, p < .001, r = .79). Additionally, members of the BR group failed MSVT at least one subtest (thus failing the MSVT) at a significantly higher rate than the UR group, $\chi^2(5, N = 50) = 33.2$, p < .001, r = .81. Classification accuracy statistics for the twelve best freestanding RVT variables are shown in Table 14.

A series of stepwise logistic regressions were conducted to identify the best predictors of invalid responding among the twelve freestanding RVT variables. VSVT Difficult Correct and VSVT Total Correct were excluded from these logistic regressions because they demonstrated perfect classification accuracy and overfit the regression models. To keep from overfitting the models, ten freestanding RVT variables were divided into four groups of similar variables:

1) VSVT Easy RT and VSVT Difficult RT;

2) VSVT Total RT and VSVT Difficult RT SD;

3) MSVT IR, MSVT DR, and MSVT CNS; and

4) MSVT PA, MSVT FR, and MSVT Fail Any Subtest.

Significant predictor variables were identified and loaded into subsequent stepwise regressions until the best predictors of the biased responder group (behind VSVT Difficult Correct and VSVT Total Correct) were identified.

Analyses indicated that VSVT Total RT and MSVT FR as a set most reliably differentiated the BR and UR groups, $\chi^2(2, N = 50) = 49.9$, p < .001. Nagelkerke's R^2 of

.84 indicated a strong relationship between prediction and grouping. Prediction success overall was 92.0% (92.3% for UR group and 91.7% for BR group). VSVT Total RT (χ^2 (1) = 4.77, p = .03) and MSVT FR (χ^2 (1) = 6.27, p = .01) both made significant contributions to group prediction. Exp(B) values indicated that each additional second in VSVT Total RT increases the likelihood of a person being an invalid responder by more than 17 times. Since higher MSVT Free Recall scores are consistent with better performance, each fewer MSVT FR percentage point increases the likelihood of a person being an invalid responder by 1.19 times. Put another way, each missed MSVT question (with 5 percentage points per question) increases the likelihood of the participant being an invalid responder by 6 times. Additional ROC analyses indicated that the combined AUC of VSVT Total RT and MSVT FR was 0.99 (95% *CI* = 0.96 to 1.00). VSVT Total RT contributed an additional 0.031 to the overall classification accuracy above and beyond MSVT FR, and MSVT FR contributed an additional 0.057 to the classification accuracy above and beyond VSVT Total RT.

Hierarchical Logistic Regressions

Once variables from the three test types were reduced to the best and most representative variables (see Figure 2), hierarchical logistic regressions were conducted to test incremental predictive ability of the BEAM above and beyond embedded RVTs (Hypothesis 2A), freestanding RVTs (Hypothesis 2B), and both embedded and freestanding RVTs as a set (Hypothesis 2C). The variables that demonstrated the best classification accuracy from their respective test type (i.e., BEAM, embedded, freestanding) were used in these analyses. Any variables that individually demonstrated perfect classification accuracy were excluded from these analyses because perfect

classification overfits regression models in SPSS. Based on concurrent validity results above, Reliable Digit Span-Revised (RDS-R) was identified as the best, most representative embedded RVT variable, as was MSVT Free Recall % Correct (FR) for freestanding RVT variables. Either Overall Saccadic or Manual Reaction Time Intra-Individual Variability (SacRT-IIV-Overall; ManRT-IIV-Overall) could be used to represent the BEAM since both variables demonstrated nearly identical classification accuracy and additional predictive value over each other.

As seen in Table 15, several models were used to test incremental predictive validity. In the first model, RDS-R was entered in block 1 and SacRT-IIV-Overall was entered in block 2. Hypothesis 2A was *confirmed* as the representative BEAM variable added significantly to the model above and beyond the representative embedded RVT variable, $\chi^2(2, N = 50) = 58.1$, $p < .001$, R^2-change = .18. In the second model, MSVT FR was entered in block 1 and SacRT-IIV-Overall was entered in Block 2. Parameter estimates could not be obtained because the model demonstrated exact classification accuracy and was overfit. As such, a separate model with ManRT-IIV-Overall was entered in block 2. The alternate representative BEAM variable added significantly to the model above and beyond the representative freestanding RVT variable[23], $\chi^2(2, N = 49) = 53.3$, $p < .001$, R^2-change = .15. Because freestanding RVTs produced two variables with perfect classification, and the BEAM cannot outperform perfection, Hypothesis 2B was *partially confirmed*.

A third model was conducted to compare embedded to freestanding to BEAM. RDS-R was entered into block 1 and MSVT FR was entered into block 2. As expected, the representative freestanding RVT variable added significantly to the model above and

[23] Without achieving exact classification accuracy

beyond the representative embedded RVT variable, $\chi^2(2, N = 50) = 52.5$, $p < .001$, R^2-change = .13. In block 2, prediction success overall was 92.0% (92.3% for UR group and 91.7% for BR group). When SacRT-IIV-Overall was entered into block 3 of this final model, the model achieved exact classification accuracy and was overfit. The same result was found when ManRT-IIV-Overall was entered into block 3 as the representative BEAM variable. While both manual and saccadic representative BEAM variables improved the overall prediction success from 92.0% to 100%, the significance of the increase in predictive value above and beyond both embedded and freestanding RVTs could not be tested in SPSS Version 20. Additionally, the BEAM's improvements to the model were obtained while using the third best freestanding RVT variable identified in the simulator study, MSVT FR. As such, Hypothesis 2C was *partially confirmed.*

SPECIFIC AIM 3

To evaluate BEAM and embedded RVT performance between simulator study participants and valid responders with a history of mild traumatic brain injury (Hypotheses 3A-3B).

At the time of the prospective, experimental study's data collection completion, data for forty-five subjects in the parent study's TBI sample had been collected. Of the thirty subjects meeting criteria for "mild TBI," two were excluded for having medical conditions the interfered with neurocognitive testing (i.e., optic nerve tumor, severe diplopia), and nine were excluded for failing one or more RVTs with 90% specificity (five failed only one RVT, four failed two or more RVTs; see Table 2 for cutoff scores). Five subjects failed CPT-II Commissions, four subjects failed CPT-II Perseverations, two

subjects failed CPT-II Hit RT SE, two subjects failed CPT-II Omissions, two subjects failed Digit Span ACSS, and two subjects failed Digit Span RDS.

Nineteen participants with a history of mild TBI met eligibility criteria for inclusion in this study. Demographic and injury information for the Unbiased Responder with a history of mild TBI (UR-mTBI) group is shown in Table 16. The UR-mTBI group did not differ significantly from either the BR or UR group on age, years of formal education, estimated premorbid intelligence, gender, or race/ethnicity. The median length of time since head injury in the UR-mTBI group was 6.9 years (IQR: 2.32 - 21.6 years). The average length of unconsciousness was 3.00 minutes (*SD* = 4.29 minutes), and the average length of post-traumatic amnesia was 18.1 minutes (*SD* = 50.8 minutes).

A series of one-way ANOVAs were used to test for BEAM and embedded RVT differences among BR, UR, and UR-mTBI groups. To reduce the probability of Type I error, only normally distributed variables with AUC $\geq$ 0.9 were considered for these analyses. Additionally, a Bonferroni correction was applied so all effects are reported at a .003 level of significance. As shown in Table 17, significant group differences were found for three embedded RVT variables: ACSS, $F(2, 66) = 38.4$, $p < .001$; RDS, $F(2, 66) = 32.0$, $p < .001$; and CPT-II Commissions, $F(2, 66) = 29.7$, $p < .001$. Significant group differences were also found for seven BEAM variables: SacRT-IIV-DC, $F(2, 66) = 36.6$, $p < .001$, SacRT-IIV-UCG, $F(2, 65) = 26.6$, $p < .001$, SacRT-IIV-Overall, $F(2, 66) = 47.0$, $p < .001$, ManRT-IIV-DC, $F(2, 65) = 54.6$, $p < .001$, ManRT-IIV-NDC, $F(2, 65) = 41.0$, $p < .001$, ManRT-IIV-MDC, $F(2, 60) = 33.6$, $p < .001$, and ManRT-IIV-Overall, $F(2, 65) = 53.0$, $p < .001$. Because Levene's test indicated unequal variances for

SacRT-IIV-MDC, $F(2, 66) = 4.37$, $p = .02$, Welch's F test was used and significant group differences were found, $F_w(2, 37.7) = 26.0$, $p < .001$.

Post-hoc Tukey HSD tests using a .05 level of significance identified BR groups as having significantly higher BEAM RT variability and CPT-II Commissions than UR and UR-mTBI groups (Cohen's *d* range: 1.35 - 2.68). Furthermore, the BR group had significantly lower WAIS-IV Digit Span ACSS and RDS scores than the UR and UR-mTBI groups (Cohen's *d* range: 1.94 - 2.23). The UR and UR-mTBI groups did not significantly differ on BEAM RT variability, ACSS, RDS, or CPT-II Commissions.

The eight non-normally distributed BEAM and embedded RVT variables with AUC ≥ 0.9 were subjected to Kruskal-Wallis tests. As shown in Table 18, there were significant group differences for four embedded RVT variables: RDS-R, $H(2) = 36.4$, $p < .001$; ARDS, $H(2) = 32.9$, $p < .001$, CPT-II Omissions, $H(2) = 32.1$, $p < .001$, and CPT-II Hit RT SE, $H(2) = 27.5$, $p < .001$. As shown in Table 18, additional Kruskal-Wallis tests identified significant group differences among four BEAM variables: SacCom%, $H(2) = 33.6$, $p < .001$; ManOm%, $H(2) = 38.1$, $p < .001$; ManRT-IIV-UCG, $H(2) = 26.5$, $p < .001$, and ManRT-IIV-UC, $H(2) = 26.8$, $p < .001$. Post-hoc Mann Whitney *U* tests revealed no significant differences between UR and UR-mTBI groups on any of the eight non-normally distributed BEAM or embedded RVT variables. However, significant differences with medium-to-large effect sizes (*r* range: .54 - .71) were found for all eight variables between the UR-mTBI and BR groups. Classification accuracy statistics with BR, UR, and UR-mTBI groups are shown in Tables 19 and 20. Based on these results, both Hypothesis 3A and 3B were *confirmed*.

CHAPTER 4: Discussion

SUMMARY AND INTEGRATION OF RESULTS

This dissertation project sought to evaluate a novel eye tracking tool's ability to detect invalid responding in neurocognitive assessment relative to existing response validity tests (RVTs) and metrics. A well-controlled, prospective, experimental simulator study with sufficient sample size (n = 50) to power analyses was conducted to identify variables from the eye tracking tool as well as embedded and freestanding RVTs that most accurately classified simulators and controls. Results obtained from the simulator study were compared to parent study data of research participants with a history of mild TBI in order to cross-validate experimental results with a clinical sample of interest.

The first aim of the project was to identify variables from the Bethesda Eye & Attention Measure (BEAM) that demonstrated the best classification accuracy among biased and unbiased responders. Twenty-five of the twenty-nine BEAM variables submitted to ROC analyses demonstrated acceptable-to-outstanding (AUC: 0.71 - 0.97) classification accuracy: ten reaction time (RT) variables, twelve reaction time intra-individual variability (RT-IIV) variables, two commission error variables (Com%), and one omission error variable (Om%). Ten RT-IIV variables, SacCom%, and ManOm% demonstrated outstanding classification accuracy (AUC ≥ 0.9). Saccadic and Manual RT-IIV-Overall demonstrated the best classification accuracy among BEAM variables (AUC = 0.97). Between-groups analyses of the twelve BEAM variables with outstanding classification accuracy identified significant ($p < .001$) differences between the biased and unbiased responding groups with large effect sizes. Compared to the unbiased responding (UR) control group, the biased responding (BR group) demonstrated slower

saccadic and manual reaction time; greater saccadic and manual reaction time intra-individual variability; more saccadic and manual commission errors; and more manual omission errors. The variable that represented compliance with test instructions failed to achieve acceptable classification accuracy. As such, Hypotheses 1C and 1D were *confirmed*, Hypotheses 1B and 1E were *partially confirmed*, and Hypothesis 1A was *not confirmed*.

The second aim of the project was to determine incremental predictive value of BEAM variables above and beyond existing indices and tests used to identify invalid responding. After taking the BEAM, simulator study participants completed the Trail Making Test (TMT) A & B, Conners' Continuous Performance Test-Second Edition (CPT-II), and the WAIS-IV Digit Span Subtest. Each of these three tests contains well-researched embedded validity metrics that can be used to detect invalid responding. Thirteen of the eighteen embedded validity metrics from these tests demonstrated acceptable-to-outstanding (AUC: 0.74 - 0.94) classification accuracy: seven from the CPT-II, four from the WAIS-IV Digit Span, and two from TMT A & B. Seven of these variables—CPT-II Commissions, Omissions, and Hit RT SE and WAIS-IV Digit Span RDS, RDS-R, ACSS, and ARDS—demonstrated outstanding classification accuracy in the simulator study sample. RDS-R was found to have the best overall classification accuracy among the embedded RVT variables (AUC = 0.94). Between-groups analyses of embedded RVT variables with outstanding classification accuracy identified significant ($p < .001$) differences between the biased and unbiased responding groups with large effect sizes.

Simulator study participants also took two freestanding RVTs—the Victoria Symptom Validity Test (VSVT) and the Medical Symptom Validity Test (MSVT)—as part of their assessment battery. All fifteen freestanding RVT variables submitted to ROC analyses demonstrated excellent-to-outstanding (AUC: 0.85 - 1.00) classification accuracy in the simulator study sample: nine from the VSVT and six from the MSVT. Twelve of these variables demonstrated outstanding classification accuracy. Two VSVT variables—Difficult Correct and Total Correct—demonstrated perfect classification accuracy and could not be submitted to the planned hierarchical logistic regression models in SPSS[24]. As such, the MSVT Free Recall (FR) % Correct variable was found to have the best overall classification accuracy among the freestanding RVT variables permissible for logistic regression analyses. Similar to BEAM and embedded RVT findings, between-groups analyses of freestanding RVT variables with outstanding classification accuracy identified significant ($p < .001$) differences between the biased and unbiased responding groups with large effect sizes.

Hierarchical logistic regressions using the best and most representative variables[25] from the BEAM (SacRT-IIV-Overall/ManRT-IIV-Overall), embedded RVTs (RDS-R), and freestanding RVTs (MSVT FR) were conducted. The best BEAM variable outperformed the best embedded RVT variable, with SacRT-IIV-Overall demonstrating incremental predictive ability above and beyond RDS-R. As such, Hypothesis 2A was *confirmed*. The best BEAM variable outperformed the third best freestanding RVT variable, with ManRT-IIV-Overall demonstrating incremental predictive value above and

[24] It is impossible to improve predictive value of a model above-and-beyond perfect group classification. While these variables were clearly the best and most representative freestanding RVT variables, they could not be used to test Hypotheses 2B and 2C.

[25] That could be submitted to the logistic regression analyses

beyond MSVT FR. As such, Hypothesis 2B was *partially confirmed*. In a third model, MSVT FR demonstrated incremental predictive ability above and beyond RDS-R, but the addition of either SacRT-IIV-Overall or ManRT-IIV-Overall to the model rendered perfect classification accuracy and overfit the model. As such, Hypothesis 2C was *partially confirmed*.

The third and final aim of the project was to compare BEAM and embedded RVT performance between simulator study participants and unbiased responders with a history of mild TBI (UR-mTBI). Variables that demonstrated outstanding classification accuracy in the simulator study were submitted to between-groups analyses and post-hoc tests. None of the nineteen BEAM and embedded RVT variables with outstanding classification accuracy differed significantly ($p < .05$) between the UR and UR-mTBI groups, but all nineteen variables differed significantly ($p < .05$) between the BR and UR-mTBI groups with medium-to-large effect sizes. The BR group consistently demonstrated greater saccadic (BEAM) and manual (BEAM, CPT-II) reaction time intra-individual variability; more saccadic (BEAM) and manual (BEAM, CPT-II) commission errors; more manual (BEAM, CPT-II) omission errors; and poorer WAIS-IV Digit Span performance than the UR-mTBI group. As such, Hypotheses 3A and 3B were *confirmed*.

EXPLANATIONS FOR FINDINGS

Overall, the prospective, experimental simulator study produced a higher than initially expected number of BEAM, embedded RVT, and freestanding RVT variables with acceptable-to-outstanding classification accuracy for invalid responding. Specifically, 53 of 62 variables (85.5%) achieved the study's baseline for inclusion in subsequent analyses. As a result of these initial findings, only variables with outstanding

classification accuracy (AUC ≥ 0.9) in the prospective, experimental design were submitted to additional analyses. Of the 62 original variables, 31 (50.0%) remained after applying this new cutoff: 12 BEAM variables, 7 embedded RVT variables, and 12 freestanding RVT variables.

Because this study was the first to evaluate the ability of a novel eye tracking tool—the BEAM—to identify invalid responding, all 29 BEAM variables derived from the custom-built scoring software were originally submitted to ROC analyses. This approach identified several saccadic and manual metrics that were able to detect invalid responding in a sample of healthy participants. The primary research question of the dissertation project ("Can eye movements be used to detect invalid responding?") was confirmed.

Interestingly, saccadic and manual variables exhibited several interesting trends in how well they classified invalid responding. In both saccadic and manual modalities, reaction time intra-individual variability (RT-IIV) metrics generally outperformed reaction time (RT) metrics in classifying invalid responding. RT-IIV metrics ranged from excellent-to-outstanding (AUC: 0.85 - 0.97), whereas RT metrics ranged from unacceptable-to-excellent (AUC: 0.60 - 0.89). Ten of the twelve BEAM variables with outstanding classification accuracy were RT-IIV metrics: four SacRT-IIV variables and six ManRT-IIV variables.

Qualitatively, Manual RT and RT-IIV metrics appeared to outperform their respective Saccadic RT and RT-IIV metrics in detecting invalid responding. ManRT-IIV variables were strong among the five non-DCR trial types (AUC: 0.92 - 0.97) and overall variability (AUC = 0.97, 95% *CI*: 0.92 – 1.00). Four of the five BEAM variables

with the highest AUC were ManRT-IIV metrics with 95% AUC confidence intervals between 0.90-1.00. These results could be explained by the greater reliability among manual metrics compared to saccadic metrics (18). While SacRT-IIV variables were relatively weaker among the non-DCR trial types (AUC: 0.85 - 0.93), SacRT-IIV-Overall (AUC = 0.97, 95% *CI*: 0.93 – 1.00) emerged as one of the two best BEAM variables for identifying invalid responding in the simulator study. It is possible that the way SacRT-IIV-Overall is calculated (i.e., average of trial type standard deviations) enhances its internal reliability and contributes to its superior performance relative to the other saccadic metrics. It is also possible that the more automatic nature of saccadic responses to visual stimuli may render them less vulnerable to invalid responding, an interpretation supported by previous research (19; 117). Based on this initial study, it is clear that both modalities offered metrics that were able to identify invalid responding with outstanding classification accuracy. Additional studies examining both the complimentary and the independent contributions that saccadic and manual metrics each provide to the evaluation of cognitive performance would help explain the findings in this study.

Interestingly, BEAM commission and omission errors manifested differently between saccadic and manual responses. Both SacCom% and ManOm% demonstrated outstanding classification accuracy. By comparison, ManCom% had AUC = 0.80 and SacOm% had AUC = 0.60. While commission errors (i.e., looking or pressing the button during the inhibition trials) appeared to have some utility in both response modalities, saccadic commissions appeared to be more sensitive to invalid responding than manual commissions.

Conversely, Manual Omission Errors were much more likely to occur in the BR group than Saccadic Omission Errors. While it was rare in both biased and unbiased groups for participants to omit any saccadic response after a target circle appeared, manual omission errors were significantly more likely to occur in the BR group. It is possible this finding was influenced by the scoring software defines omission errors (i.e., no response in 1000ms window on non-inhibition trials). Any button press occurring more than 1000ms after the target circle appeared would be counted as a manual omission, just as if no button press occurred at all. Given that manual reaction time is consistently longer than saccadic reaction time on the BEAM (18), omission errors may have been more likely to be registered in the manual modality. Still, it was surprising to see such a disparity between modalities, with ManOm% classification accuracy being in the outstanding range and SacOm% being in the unacceptable range.

Several findings among the embedded and freestanding RVTs met or exceeded expectations of their classification accuracy. Consistent with previous literature (42; 234; 269), the WAIS-IV Digit Span and CPT-II demonstrated better capability to detect invalid responding than the Trail Making Test A & B. Not surprisingly, all fifteen variables from the VSVT and MSVT—tests that were designed to detect invalid responding—demonstrated excellent-to-outstanding classification accuracy. Consistent with the literature (248), the VSVT Difficult Correct (and by extension, Total Correct) variable obtained the best overall classification accuracy in the simulator study.

Classification accuracy statistics for BEAM, embedded RVT, and freestanding RVT cutoff scores were strong across the three test types (i.e., BEAM, embedded RVTs, freestanding RVTs). Using Larrabee's recommended 0.85 minimum specificity level for

detecting invalid responding (153; 154), BEAM variables demonstrated sensitivity levels ranging from 0.74 - .0.96. Using the same specificity criterion, embedded RVT variable sensitivities ranged from 0.83 - 0.96 and freestanding RVT variable sensitivities ranged from 0.75 - 1.00. Collectively, all 31 variables with outstanding classification accuracy among the three test types were able to detect roughly three quarters of the biased responders while maintaining 85% specificity. It should also be noted that the base rate of invalid responding in the simulator study was 48% (24 out of 50 were biased responders). While this experimental base rate approximates actual base rates in real-world clinical contexts (see Base Rates section above), smaller base rates (in the 10-20% range) would render smaller PPVs and higher NPVs than reported in Table 7.

As intended, the well-controlled, prospective experimental design rendered significant differences in neurocognitive assessment performance between the biased and unbiased responders. The randomized, permuted block assignment with a substantial sample size (n = 50) rendered demographically similar groups. Examiner scoring bias was mitigated by blinding and computer-automated, objective data collection. Consistent with Sollman and Berry's (248) meta-analytic findings, providing participants with a warning to fake believably achieved a large performance disparity between biased and unbiased responders. Collectively, these findings suggest the research design effectively isolated the "invalid responding" construct, enhancing the internal validity of the simulator study results.

As internal validity increases, external validity usually decreases. Not surprisingly, the embedded and freestanding RVT variables' classification accuracies were generally higher in the simulator study than previous studies using clinical

populations. Multiple studies have reported CPT-II Commissions, Omissions, and Hit RT SE variables having point AUCs in the acceptable range ($0.7 \leq AUC < 0.8$) among clinical referrals for neuropsychological testing (42; 149; 198). In a retrospective study of veterans referred for neuropsychological assessment, Young and colleagues (284) reported 95% AUC confidence intervals in the unacceptable-to-acceptable ($0.6 \leq AUC < 0.8$) range for WAIS-IV Digit Span ACSS, RDS, and RDS-R variables. Similarly, Busse and Whiteside (42) reported TMT A having unacceptable (AUC = 0.62) classification accuracy while TMT B demonstrated acceptable (AUC = 0.76) classification accuracy among a sample of 413 consecutive neuropsychological referrals.

The relatively higher AUCs in this project's simulator study may be explained by the non-clinical subject pool and well-controlled experimental manipulation of invalid responding. Consistent with the UR-mTBI group's somewhat poorer neurocognitive performance relative to the UR group, the AUCs for BEAM and embedded RVT variables were somewhat lower when comparing BR and UR-mTBI groups. Classification accuracy statistics for BEAM, embedded RVT, and freestanding RVT variables obtained in the simulator study would likely be lower in more heterogeneous, clinical samples.

The simulator study's statistical methodology utilized a two-prong approach to identifying the best detectors of invalid responding among each test type. First, ROC analyses were used in an exploratory manner to reduce a large number of variables into a more manageable subset. These ROC curves provided a cross-sectional snapshot of overall performance. Next, stepwise logistic regressions were performed on this reduced set of variables to more systematically compare their group classification prediction

capabilities. The logistic regressions were able to distill the maximum benefit from multiple scores and indices, making the technique particularly useful for identifying new embedded response validity indices in the BEAM (181). The logistic regression models accounted for fluctuating responding behavior throughout the battery by including variables from multiple measures obtained at several time points (36; 234). Results from stepwise regression analyses were compared with joint variable ROC analyses (i.e., CPT-II Commissions + WAIS-IV Digit Span RDS-R) to identify the most representative variables from each test type.

The ROC and stepwise logistic regression analyses independently identified the same variables from each test type as being the best predictors of invalid responding: SacRT-IIV-Overall (AUC = 0.97) and ManRT-IIV-Overall (AUC = 0.97) from the BEAM; WAIS-IV Digit Span RDS-R (AUC = 0.94) from embedded RVTs; and VSVT Difficult Correct (AUC = 1.00), VSVT Total Correct (AUC = 1.00), and MSVT FR (AUC = 0.96) from freestanding RVTs. The best variables identified in Aim 1 of this study, as shown in Table 4, were the same variables that were identified by a series of stepwise logistic regression models. These convergent findings from separate statistical methodologies enhanced the confidence that the variables submitted to hierarchical regression analyses would best represent their respective test types. Of note, both a saccadic and manual metric from the BEAM could independently be used to represent the BEAM.

Earlier in the study, eye movement metrics were shown to be useful to detecting invalid responding; hierarchical logistic regressions were then used to see how useful they were when compared to existing measures (the second part of this study's primary

research question). Analyses revealed that the BEAM demonstrated considerable predictive ability of invalid responding above and beyond embedded RVTs. While two freestanding RVT variables with perfect classification accuracy outperformed the BEAM in terms of AUC, it could not be determined if VSVT Difficult/Total Items Correct significantly improved group prediction above and beyond the BEAM. It was encouraging nonetheless to find that the BEAM demonstrated moderate predictive ability of invalid responding above and beyond all other freestanding RVT variables, including all MSVT variables.

Consistent with the literature (179), freestanding RVTs (not including VSVT Difficult/Total Correct variables) demonstrated incremental predictive value above and beyond embedded RVTs in the simulator study. The addition of a either a saccadic or manual BEAM variable in this Embedded RVT + Freestanding RVT model achieved perfect classification accuracy. While this finding indicates an objective improvement in group prediction, limitations of the statistical methodology preclude understanding as to whether the BEAM significantly added to the Embedded RVT + Freestanding RVT model.

Collectively, the prospective simulator study results provide compelling evidence that the BEAM's saccadic and manual metrics may serve as powerful tools for detecting invalid responding. The results suggest that the BEAM's saccadic and manual metrics could be used to detect invalid responding similar to the well-researched embedded RVT measures in the CPT-II, TMT, and WAIS-IV. The simulator study results also provide preliminary evidence that BEAM metrics may perform comparably to the MSVT and VSVT at detecting invalid responding.

To enhance generalizability of the simulator study's findings, the experimental study results needed to be compared to a group with a known clinical condition. Mild TBI populations are known for having high rates of invalid responding (44; 100; 183), and the parent study enabled BEAM and embedded RVT variables to be compared to a group with this condition. Conservative screening methods increased confidence that the mild TBI participants were performing at capacity levels (i.e., valid responding). Because the clinical comparison group did not differ demographically from the UR or BR groups, group analyses were able to be conducted without controlling for demographic variables and limiting statistical power. While freestanding RVT comparison data were unavailable, analyses were able to be conducted with the 19 BEAM and embedded RVT variables that demonstrated outstanding classification accuracy.

While the UR-mTBI group had somewhat poorer BEAM and embedded RVT scores than the UR group, the performance differences were not significant. These null findings could be explained by the UR-mTBI group's median time since injury of 6.9 years and an above-average premorbid intelligence. It is possible that any effects of neurological injury at time of testing were attenuated by time, cognitive ability, or behavioral factors that limited the ability of conventional neuropsychological testing to identify significant impairment (72). It is also possible that this study was underpowered to detect significant differences in neurocognitive performance among persons with mild TBI and healthy controls. Nonetheless, the UR-mTBI group served a useful purpose in this study by providing a clinical comparison group to healthy groups of biased and unbiased responders. In contrast to the UR/UR-mTBI group comparisons, the BR group

performed significantly worse than the UR-mTBI group on all 19 BEAM, CPT-II, and WAIS-IV Digit Span variables with medium-to-large effect sizes.

Further analyses of cutoff scores in the clinical sample enhanced the generalizability of the simulator study's findings. BEAM variables generally demonstrated excellent sensitivity (0.61 - 0.92) to invalid responding while maintaining 0.85 or greater specificity in the UR-mTBI group. This clinical sensitivity range for BEAM variables is comparable to the 0.74 - 0.96 range obtained in the simulator study, suggesting that the BEAM's invalid responding detection performance may not be significantly degraded in validly responding mild TBI populations. By contrast, the WAIS-IV Digit Span and CPT-II variables demonstrated sensitivity levels ranging from 0.50 - 0.75 with specificity levels of 0.85 or above in the UR-mTBI group. This clinical sensitivity range for embedded RVT variables is much lower than the 0.83 - 0.96 sensitivity range obtained in the simulator study and more consistent with previous embedded RVT research with clinical samples (42; 149; 198; 284).

Despite not having significant group differences between the UR and UR-mTBI group on BEAM, CPT-II, and WAIS-IV Digit Span performance, it was not surprising to see cutoff scores with specificity on or about 0.85 in the control group result in lower specificity in the clinical group. This lowered clinical specificity resulted in higher numbers of false positives across several BEAM and embedded RVT variables, including ones identified as being the best and most representative variables in the simulator study. SacRT-IIV-Overall scores of 0.108 sec or greater had 88% sensitivity to invalid responding and 85% specificity in the UR group, but the same cutoff score demonstrated 63% specificity in the UR-mTBI group. ManRT-IIV-Overall scores of 0.092 sec or

greater had 96% sensitivity and 85% specificity in the UR group, but also demonstrated 63% specificity in the UR-mTBI group. WAIS-IV Digit Span RDS-R scores of 15 or less had 92% sensitivity and 88% specificity in the UR group, but demonstrated 63% specificity in the UR-mTBI group.

Given that UR-mTBI group cutoff scores were generally less specific to invalid responding than those in the UR group, it was surprising to find that some BEAM and embedded RVT variables produced cutoff scores with similar specificities among clinical and control groups. Five variables with outstanding classification accuracy in the simulator study had cutoff scores with similar ~15% false positive rates among the clinical and healthy control groups: SacRT-IIV-DC (≥0.104 sec cutoff score: 79% sensitivity to BR, 84% UR-mTBI specificity, 85% UR specificity), ManRT-IIV-MDC (≥0.111 sec cutoff score: 83% sensitivity to BR, 84% UR-mTBI specificity, 85% UR specificity), and ManOm% (≥1.0% cutoff: 92% sensitivity to BR, 89% UR-mTBI specificity, 88% UR specificity), WAIS-IV Digit Span RDS (≤9 cutoff: 92% sensitivity to BR, 84% UR-mTBI specificity, 88% UR specificity), and CPT-II Omissions (≥3 raw cutoff: 71% sensitivity to BR, 84% UR-mTBI specificity, 85% UR specificity).

Several BEAM and embedded RVT variable cutoff scores achieved similar ~5% false positive rates between the control and clinical groups. SacRT-IIV-DC (≥0.125 sec cutoff score: 79% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity) and ManRT-IIV-DC (≥0.123 sec cutoff score: 83% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity) demonstrated good sensitivity to invalid responding with high clinical and control group specificity. One representative BEAM variable, ManRT-IIV-Overall, performed much better across clinical and control groups when using a

cutoff score with higher specificity (≥0.118 sec cutoff score: 78% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity). ManOm% (≥8.2% cutoff score: 54% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity) and ManRT-IIV-UC (≥0.122 sec cutoff score: 57% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity) demonstrated relatively lower sensitivity than other BEAM variables with similar ~5% false positive rates among clinical and control groups. Among embedded RVT variables, WAIS-IV Digit Span ACSS (≤8 cutoff score: 71% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity), RDS (≤8 cutoff score: 67% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity), and RDS-R (≤13 sec cutoff score: 75% sensitivity to BR, 95% UR-mTBI specificity, 96% UR specificity) demonstrated varying sensitivity while maintaining similar ~5% false positive rates among clinical and control groups.

Among BEAM variables, SacRT-IIV-DC, ManRT-IIV-DC, ManRT-IIV-UC, ManRT-IIV-Overall, and ManOm% demonstrated similar ~5% false positive rates among clinical and control groups, as did SacRT-IIV-DC, ManRT-IIV-MDC, and ManOm% at the ~15% level. Among embedded RVT variables, CPT-II Omissions, WAIS-IV Digit Span ACSS, and WAIS-IV Digit Span RDS-R demonstrated similar ~5% false positive rates among clinical and control groups, as did CPT-II Omissions and WAIS-IV Digit Span RDS at the 15% level. SacRT-IIV-DC, ManOm%, and WAIS-IV Digit Span RDS variables performed similarly in clinical and control groups at both ~5% and ~15% false positive levels. Several BEAM variables representing divergent metrics (i.e., saccadic and manual reaction time intra-individual variability, manual omission errors) demonstrated high sensitivity (0.78 - 0.92) to invalid responding while minimizing false

positives (~15%) in both clinical and control groups (See Figures 3-5). These results indicate that certain BEAM variables can be sensitive and specific to invalid responding in mild TBI populations.

The collective results from the simulator study and initial clinical validation provide preliminary evidence that the BEAM could be used in neuropsychological evaluations in manners similar to other performance measures with well-researched embedded RVTs. If efficiency is a priority, future studies may support using the BEAM in lieu of a lengthier freestanding RVT. At 12 minutes, the BEAM's total administration time (including calibration, instructions, practice, and actual measure) is faster than the CPT-II, MSVT, or VSVT. While the present costs of using licensed copies of the CPT-II, Digit Span, MSVT, and VSVT in clinical and research settings may be lower than the cost of the prototype eye-tracking system used in this study to administer the BEAM, future advances in technology will likely drive down the cost of eye-tracking hardware. If saccadic metrics prove to be clinically useful in the assessment of cognitive performance and response validity, the benefits of using eye-tracking may justify the costs.

LIMITATIONS

As with all studies, the internal and external validity of this dissertation project was limited by study design methodology, statistical analyses, and the data that was available to answer the pertinent research questions. The simulator study, as tightly controlled as it was, would have benefitted from having a larger and more diverse sample. Because demographic factors can influence the prevalence and type of invalid responding behavior (10), the above average premorbid intelligence and years of

education in all three of this study's groups may limit the generalizability of the study's findings to groups of lesser intellectual capacity. While it is entirely possible that the BR group's above-average intelligence may have produced more sophisticated (and harder-to-detect) faking than a sample with normal or below-average intelligence, this question could not be analyzed without a larger sample of participants.

Despite blinding the examiners to group assignment, it is possible that examiners may have exhibited scoring biases during data collection. Specifically, they may have biased scores lower or higher, depending on the group to which they believed the participant had been assigned. Because every test administered was either designed to identify (or could be used to identify) invalid responding, it could be expected that examiners would correctly guess group assignment for participants who appeared to perform poorly on testing. Not surprisingly, post-examination feedback data indicated that the study's examiners correctly guessed a participant group assignment 96% of the time (48/50 correct) with an average confidence of 4.73 (*SD* = 0.62) on a scale from 1 to 5 (with 5 being the most confident). In some research or clinical settings, this result could pose a limitation to the internal validity of the results; however, the risk of examiner bias in this study was largely mitigated by the automated scoring software in the BEAM, VSVT, MSVT, and CPT-II, and the objective scoring criteria from the other tests in the battery. Further research exploring examiners' behavioral observations and participants' response strategies during response validity assessment would provide useful complimentary information to the quantitative data.

Due to the differential test length and multi-part nature of some tests (i.e., 10 minute delays before "Part 2" of a test) in the neurocognitive battery, testing order

needed to remain fixed for all simulator study participants. Additionally, the UR-mTBI participants took a different battery than the UR and BR groups, one which included several additional tests and did not include the VSVT or MSVT. As such, it is possible test order effects may have influenced performance. These order effects were somewhat mitigated by comparing performance on similar tests from two separate neurocognitive batteries (UR & BR to UR-mTBI), each with a unique, fixed test order. Nonsignificant findings between UR and UR-mTBI groups coupled with significant findings between BR and UR-mTBI groups suggest that test order effects, if present, were negligible in comparison to valid or invalid responding.

The aim of the study that sought to evaluate incremental predictive value of the BEAM was limited by statistical methodology and analysis software. Specifically, logistic regression models in SPSS Version 20 become overfit when they obtain perfect classification accuracy. Due to the large group differences in the simulator study, perfect classification accuracy was quickly obtained with only minimal combinations of variables. It also forced this author to utilize a single, representative variable for each test type rather than including several variables from the BEAM, embedded RVTs, and freestanding RVTs. While it was not possible to evaluate every metric, the obtained results were qualitatively consistent with the testable ones and qualitatively interpretable. The logistic regression models provided useful adjunctive information to the ROC analyses in Aim 1.

Ideally, this project would have had the time and resources to collect more data from persons with a history of mild TBI. As stated previously, a true combined groups study would have incorporated at least four groups to maximize internal and external

validity: a biased experimental group (without a clinical condition), an unbiased experimental group, and biased clinical group, and an unbiased clinical group (218; 249). While this study included three of these groups, it was not possible to incorporate a biased clinical group of participants with a history of mild TBI. As such, our understanding of how a clinical group would have demonstrated invalid responding on the BEAM in comparison to other groups is limited.

Other studies using known groups in mild TBI samples have traditionally used a "gold-standard" criterion such as freestanding RVT like the WMT (149; 151) or TOMM (42) failure to split groups into biased and unbiased responders. In the absence of freestanding RVT data in the parent study, multiple embedded RVT indices were used to rigorously screen for invalid responding. Because embedded RVTs are generally less sensitive than freestanding RVTs (179), it was determined that a single embedded validity index failure with 90% specificity disqualified a subject from the UR-mTBI group. While this method may have reasonably precluded 10% of the valid responders with a history of mild TBI from our sample, it was determined that the criteria for inclusion in the UR-mTBI group needed to be as stringent as possible to enhance confidence that the "valid responders" were responding validly.

Finally, it should be noted that the construct being examined in this dissertation project (i.e., "true" invalid responding) is rarely observed with certainty or experimentally induced. However, behavior that *approximates* true invalid responding can be observed and experimentally induced with the proper research design, and this approximate behavior can nonetheless serve as a useful standard of reference for future clinical judgments. Obtaining cutoff scores from controlled, experimental studies of non-

clinical groups isolates the effects of the experimentally manipulated "invalid responding" construct. Results from these studies can then be applied to clinical groups to examine overlap between invalid responding performance and capacity performance impacted by some clinical condition. By following research guidelines established by leaders in the field (34; 154; 218), this project enhanced its contributions to the field of response validity research.

FUTURE DIRECTIONS

This study was the first to examine the BEAM's ability to identify invalid responding in a neurocognitive assessment, and subsequent studies are needed to verify and replicate this study's findings in clinical and non-clinical groups with known groups of valid and invalid responders. The results of this study raise several important questions for future studies. Specifically, which BEAM variables will perform the best towards discriminating valid from invalid responders with varying neurocognitive disorders? How will the BEAM's ability to detect invalid responding compare with existing embedded and freestanding RVTs in subsequent studies? How will BEAM performance manifest in research, clinical, and disability evaluation contexts within the DoD/VA? How will the BEAM's performance be impacted by fatigue, depression, anxiety, chronic pain, and other factors that influence test-taking ability?

The BEAM's concurrent, dual-response modality (where participants look at visual stimuli *and* press a button in response to a single target) may elicit wide-ranging invalid responding strategies, and these strategies should be examined in greater detail. The wide net of conceptually unique BEAM variables that obtained outstanding classification accuracy in this study—saccadic and manual metrics, commission and

omission errors—suggests that a diagnostic algorithm could be developed for invalid responding. This algorithm could incorporate performance across BEAM variables into Slick et al.-like (244-246) criteria for possible, probable, and definite invalid responding.

Additional studies with clinical populations are needed to assess the BEAM's "real-world" classification accuracy. While this study showed promise that the BEAM could minimize false positives in a clinical sample, additional studies of valid and invalid responders with head injuries are needed. It is currently unknown how invalid responders with a history of mild TBI would perform on the BEAM. Other studies of persons with clinical conditions of interest should compare BEAM performance with embedded and freestanding RVTs in order to cross-validate incremental predictive value of the BEAM above-and-beyond existing measures.

CONCLUSION

The results of this combined study strongly suggest that eye movements can be used to detect invalid responding at or above the capabilities of several existing response validity tests. The Bethesda Eye & Attention Measure (BEAM), a novel eye-tracking tool, appears uniquely capable of serving as a neurocognitive assessment tool with multiple, unique embedded validity indices. The BEAM performed favorably when compared to well-validated embedded and freestanding RVTs—including the CPT-II, WAIS-IV Digit Span, Trail Making Test A & B, and the MSVT.

This project's findings support the general trend of using continuous performance tests and their associated metrics (e.g., RT, RT variability, omissions, commissions) to detect invalid responding (42; 113; 118; 149; 169; 198; 211; 253; 264; 282). The study adds to the extant response validity literature by demonstrating that saccadic performance

in a continuous performance test may be used to detect invalid responding. This project also provides evidence that concurrent, dual-modality tests such as the BEAM (i.e., saccadic and manual responses) may provide enhanced capabilities of detecting invalid responding above and beyond single-modality tests (i.e., manual-only responses) in clinical and non-clinical samples.

The generalizability of the simulator study results were greatly enhanced with the addition of a clinical comparison group of valid responders with a history of mild TBI. In addition to its strong performance in tightly controlled experimental conditions, the BEAM demonstrated preliminary evidence supporting its clinical utility. Results from the clinical group analyses suggest that the BEAM could potentially identify invalid responding in larger, more diverse mild TBI populations while minimizing false positives. Additional research should evaluate the BEAM's ability to identify invalid responding among other neurological conditions such as ADHD, HIV-associated neurocognitive disorder (HAND), and moderate-to-severe TBI. While future studies are needed to cross-validate this project's findings in larger, more heterogeneous groups of persons with and without neurological conditions, this project's collective findings provide strong initial evidence that the BEAM can be used to detect invalid responding in neurocognitive assessment.

APPENDIX A: TABLES

Table 1: Contingency table for response validity test outcomes in a TBI population

	Actual Diagnostic Condition	
Test Result	Invalid Performance-Yes	TBI-No
Invalid Performance-Yes	TP (true positive)	FP (false positive = α =Type I error)
Invalid Performance-No	FN (false negative = β =Type II error)	TN (true negative)

Table 2: Embedded response validity test cutoff scores used for UR-mTBI group

Measure	**Metric**	**Cutoff**
WAIS-IV Digit Span	Reliable Digit Span (RDS)	≤ 7
WAIS-IV Digit Span	Age Corrected Scaled Score (ACSS)	≤ 7
CPT-II	Omissions	≥ 12 raw
CPT-II	Commissions	≥ 22 raw
CPT-II	Hit RT SE	≥ 14 raw
CPT-II	Perseverations	≥ 2 raw
TMT A	Completion Time	≥ 63 sec
TMT B	Completion Time	≥ 200 sec

Table 3:Demographic characteristics of simulator study groups

	Biased Responders (n=24)	**Unbiased Responders (n=26)**	***p***
Mean age in years (SD)	28.6 (8.9)	28.4 (10.5)	.92
Mean years of education (SD)	16.9 (2.0)	16.7 (1.7)	.67
Estimated premorbid intelligence (SD)	117 (9.3)	116 (7.8)	.78
Gender			.54
Male	9 (37.5)	12 (46.2)	
Female	15 (62.5)	14 (53.8)	
Race/ethnicity			.13
Caucasian	16 (66.7)	21 (80.8)	
African-American	1 (4.2)	4 (15.4)	
Hispanic	1 (4.2)	0 (0.0)	
Asian	5 (20.8)	1 (3.8)	
Other	1 (4.2)	0 (0.0)	
Knowledge of head injury sequelae (SD)	9.48 (2.63)	9.58 (2.23)	.89

Note. *p* values reflect results of t-tests or chi-square analyses

Table 4: ROC analyses for BEAM variables in simulator study (BR vs. UR groups)

Variable	Positive (BR)	Negative (UR)	AUC	SE	95% Low	95% Hi	*p*
Saccadic RT-IIV-Overall	24	26	0.97	0.02	0.93	1.00	<.001
Manual RT-IIV-Overall	23	26	0.97	0.02	0.92	1.00	<.001
Manual RT-IIV-DC	23	26	0.97	0.02	0.92	1.00	<.001
Manual RT-IIV-NDC	23	26	0.96	0.03	0.90	1.00	<.001
Manual RT-IIV-MDC	18	26	0.96	0.03	0.90	1.00	<.001
Manual Omission Error %	24	26	0.94	0.04	0.87	1.00	<.001
Saccadic Commission Error %	24	26	0.94	0.03	0.87	1.00	<.001
Saccadic RT-IIV-DC	24	26	0.93	0.04	0.86	1.00	<.001
Manual RT-IIV-UCG	21	26	0.93	0.05	0.84	1.00	<.001
Saccadic RT-IIV-MDC	24	26	0.92	0.05	0.83	1.00	<.001
Manual RT-IIV-UC	23	26	0.92	0.04	0.84	0.99	<.001
Saccadic RT-IIV-UCG	23	26	0.90	0.05	0.80	1.00	<.001
Manual RT-Overall	23	26	0.89	0.05	0.80	0.99	<.001
Manual RT-UCG	21	26	0.89	0.05	0.78	0.99	<.001
Manual RT-NDC	23	26	0.88	0.05	0.78	0.98	<.001
Manual RT-MDC	18	26	0.86	0.07	0.74	0.99	<.001
Saccadic RT-IIV-UC	24	26	0.85	0.05	0.75	0.96	<.001
Saccadic RT-IIV-NDC	24	26	0.85	0.06	0.74	0.96	<.001
Manual RT-DC	23	26	0.84	0.06	0.73	0.96	<.001
Manual RT-UC	23	26	0.83	0.06	0.70	0.95	<.001
Saccadic RT-MDC	24	26	0.81	0.06	0.69	0.93	<.001
Manual Commission Error %	24	26	0.80	0.07	0.66	0.94	<.001
Saccadic RT-DC	24	26	0.77	0.07	0.65	0.90	.001
Saccadic RT-Overall	24	26	0.74	0.07	0.60	0.88	.004
Saccadic RT-NDC	24	26	0.71	0.07	0.57	0.86	.01
# Invalid Initial Fixations	24	26	0.69	0.08	0.53	0.84	.02
Saccadic RT-UCG	23	26	0.61	0.09	0.44	0.78	.19
Saccadic RT-UC	24	26	0.60	0.09	0.42	0.77	.25
Saccadic Omission Error %	24	26	0.60	0.08	0.44	0.75	.25

Note. UR = unbiased responders; BR = biased responders; RT = reaction time; IIV = intra-individual variability; DC = directional cue; NDC = nondirectional cue; MDC = misdirectional cue; UCG = uncued with gap; UC = uncued; DCR = directional cue-red arrow.

Table 5: Non-normally distributed BEAM variables with outstanding classification accuracy

	Unbiased responders			Biased responders				
Measures	*n*	*Mdn*	*IQR*	*n*	*Mdn*	*IQR*	*p*	*r*
Saccadic Commission Error %	26	7.29	2.42 - 17.6	24	65.5	45.8 - 79.9	<.001	.75
Manual Omission Error %	26	0.00	0.00 - 0.63	24	9.65	3.91 - 26.11	<.001	.78
Manual RT-IIV-UCG (sec)	26	0.072	0.059 - 0.086	21	0.128	0.112 - 0.128	<.001	.72
Manual RT-IIV-UC (sec)	26	0.085	0.075 - 0.095	23	0.130	0.103 - 0.150	<.001	.71

Note. Mann-Whitney *U* effect size "r" was calculated by dividing Z-score by the square root of *N*. RT = reaction time; IIV = intra-individual variability; UCG = uncued with gap; UC = uncued.

Table 6: Normally distributed BEAM variables with outstanding classification accuracy

	Unbiased responders			Biased responders				
Measures	*n*	*M*	*SD*	*n*	*M*	*SD*	*p*	Cohen's *d*
Saccadic RT-IIV-DC (sec)	26	0.078	0.024	24	0.139	0.031	<.001	2.20
Saccadic RT-IIV-MDC (sec)	26	0.080	0.023	24	0.160	0.050	<.001	2.06
Saccadic RT-IIV-UCG (sec)	26	0.076	0.021	23	0.137	0.038	<.001	1.99
Saccadic RT-IIV-Overall (sec)	26	0.082	0.018	24	0.142	0.026	<.001	2.68
Manual RT-IIV-DC (sec)	26	0.076	0.021	23	0.151	0.030	<.001	2.90
Manual RT-IIV-NDC (sec)	26	0.065	0.022	23	0.128	0.027	<.001	2.56
Manual RT-IIV-MDC (sec)	26	0.077	0.025	18	0.144	0.034	<.001	2.25
Manual RT-IIV-Overall (sec)	26	0.076	0.018	23	0.134	0.022	<.001	2.89

Note. RT = reaction time; IIV = intra-individual variability; DC = directional cue; NDC = nondirectional cue; MDC = misdirectional cue; UCG = uncued with gap; UC = uncued.

Table 7: Cumulative percentages of persons with scores above the indicated cutoff on selected BEAM variables

Variable	*Cutoff*	*% BR*[a]	*% UR*[b]	*Hit Rate*	*LR+*	*LR-*	*PPV*	*NPV*
SacRT-IIV-DC (sec)	≥0.142	50					100	68
	0.137	54					100	70
	0.131	58					100	72
	0.130	63					100	74
	0.129	67					100	76
	0.128	71					100	79
	0.127	71	4	84	18.4	7.6	94	78
	0.126	75	4	86	19.5	6.5	95	81
	0.125	79	4	88	20.6	5.4	95	83
	0.119	79	8	86	10.3	2.7	90	83
	0.113	79	12	84	6.9	1.8	86	82
	0.109	79	15	82	5.2	1.4	83	81
	0.104	83	15	84	5.4	1.1	83	85
SacRT-IIV-MDC (sec)	≥0.144	63					100	74
	0.142	67					100	76
	0.138	71					100	79
	0.132	75					100	81
	0.129	79					100	84
	0.126	83					100	87
	0.123	83	4	90	21.7	4.3	95	86
	0.122	88	4	92	22.8	3.3	95	89
	0.119	88	8	90	11.4	1.6	91	89
	0.115	88	12	88	7.6	1.1	88	88
	0.112	88	15	86	5.7	0.8	84	88
SacRT-IIV-UCG (sec)	≥0.137	61					100	74
	0.136	65					100	76
	0.131	70					100	79
	0.125	74					100	81
	0.122	78					100	84
	0.117	78	4	86	20.4	5.7	95	83
	0.109	78	8	84	10.2	2.8	90	83
	0.102	78	12	82	6.8	1.9	86	82
	0.099	78	15	80	5.1	1.4	82	81
	0.097	83	15	82	5.4	1.1	83	85
	0.095	87	15	84	5.7	0.9	83	88
SacRT-IIV-Overall (sec)	≥0.133	67					100	76
	0.131	71					100	79
	0.126	75					100	81
	0.121	79					100	84
	0.118	79	4	88	20.6	5.4	95	83
	0.117	83	4	90	21.7	4.3	95	86
	0.116	83	4	88	10.8	2.2	91	86
	0.115	88	8	90	11.4	1.6	91	89
	0.112	88	12	88	7.6	1.1	88	88
	0.108	88	15	86	5.7	0.8	84	88
	0.106	92	15	88	6.0	0.5	85	92
	0.102	96	15	90	6.2	0.3	85	96
ManRT-IIV-DC (sec)	≥0.147	61					100	74
	0.145	65					100	76
	0.142	70					100	79
	0.137	74					100	81
	0.130	78					100	84
	0.127	83					100	87
	0.123	83	4	88	21.5	4.5	95	86
	0.120	83	8	86	10.7	2.3	90	86
	0.117	87	8	88	11.3	1.7	91	89
	0.105	91	8	90	11.9	1.1	91	92

Variable	*Cutoff*	*% BR*[a]	*% UR*[b]	*Hit Rate*	*LR+*	*LR-*	*PPV*	*NPV*
	0.094	91	12	88	7.9	0.8	88	92
	0.093	91	15	86	5.9	0.6	84	92
ManRT-IIV-NDC (sec)	≥0.138	43					100	67
	0.134	48					100	68
	0.129	52					100	70
	0.126	57					100	72
	0.119	61					100	74
	0.113	65					100	76
	0.111	65	4	80	17.0	9.0	94	76
	0.110	70	4	82	18.1	7.9	94	78
	0.109	74	4	84	19.2	6.8	94	81
	0.108	78	4	86	20.4	5.7	95	83
	0.106	83	4	88	21.5	4.5	95	86
	0.104	87	4	90	22.6	3.4	95	89
	0.101	91	4	92	23.7	2.3	95	93
	0.099	91	8	90	11.9	1.1	91	92
	0.094	91	15	86	5.9	0.6	84	92
	0.092	96	15	88	6.2	0.3	85	96
ManRT-IIV-MDC (sec)	≥0.141	44					100	72
	0.136	50					100	74
	0.132	56					100	76
	0.127	61					100	79
	0.125	67					100	81
	0.123	72					100	84
	0.120	72	4	76	18.8	7.2	93	83
	0.118	72	8	74	9.4	3.6	87	83
	0.116	78	8	76	10.1	2.9	88	86
	0.114	78	12	74	6.7	1.9	82	85
	0.111	83	15	74	5.4	1.1	79	88
ManRT-IIV-UCG (sec)	≥0.140	19					100	60
	0.138	24					100	62
	0.135	29					100	63
	0.132	33					100	65
	0.131	38					100	67
	0.130	43					100	68
	0.129	43	4	68	11.1	14.9	90	68
	0.128	48	4	70	12.4	13.6	91	69
	0.127	52	4	72	13.6	12.4	92	71
	0.126	57	4	74	14.9	11.1	92	74
	0.125	62	4	76	16.1	9.9	93	76
	0.123	67	4	78	17.3	8.7	93	78
	0.118	71	4	80	18.6	7.4	94	81
	0.115	76	4	82	19.8	6.2	94	83
	0.112	76	8	80	9.9	3.1	89	83
	0.107	81	8	82	10.5	2.5	89	86
	0.103	86	8	84	11.1	1.9	90	89
	0.100	86	12	82	7.4	1.2	86	88
	0.097	86	15	80	5.6	0.9	82	88
	0.096	90	15	82	5.9	0.6	83	92
	0.095	95	15	84	6.2	0.3	83	96
ManRT-IIV-UC (sec)	≥0.154	4					100	54
	0.153	9					100	55
	0.151	13					100	57
	0.151	17					100	58
	0.150	22					100	59
	0.146	26					100	60
	0.140	26	4	62	6.8	19.2	86	60
	0.136	30	4	64	7.9	18.1	88	61
	0.134	35	4	66	9.0	17.0	89	63
	0.132	39	4	68	10.2	15.8	90	64

Variable	*Cutoff*	*% BR*[a]	*% UR*[b]	*Hit Rate*	*LR+*	*LR-*	*PPV*	*NPV*
	0.131	43	4	70	11.3	14.7	91	66
	0.130	48	4	72	12.4	13.6	92	68
	0.126	52	4	74	13.6	12.4	92	69
	0.122	57	4	76	14.7	11.3	93	71
	0.119	61	4	78	15.8	10.2	93	74
	0.117	61	8	76	7.9	5.1	88	73
	0.111	65	8	78	8.5	4.5	88	75
	0.105	65	12	76	5.7	3.0	83	74
	0.104	74	15	78	4.8	1.7	81	79
ManRT-IIV-Overall (sec)	≥0.136	57					100	72
	0.135	61					100	74
	0.132	65					100	76
	0.128	70					100	79
	0.127	74					100	81
	0.122	78					100	84
	0.117	78	4	86	20.4	5.7	95	83
	0.117	78	8	84	10.2	2.8	90	83
	0.111	83	8	86	10.7	2.3	90	86
	0.106	87	8	88	11.3	1.7	91	89
	0.099	91	8	90	11.9	1.1	91	92
	0.094	96	8	92	12.4	0.6	92	96
	0.093	96	12	90	8.3	0.4	88	96
	0.092	96	15	88	6.2	0.3	85	96
Saccadic Commissions	≥57.9	58					100	72
(% of DCR trials)	55.4	63					100	74
	53.3	67					100	76
	52.6	71					100	79
	47.9	75					100	81
	43.5	79					100	84
	43.1	83	4	90	21.7	4.3	95	86
	41.8	83	8	88	10.8	2.2	91	86
	32.9	83	12	86	7.2	1.4	87	85
	24.0	83	15	84	5.4	1.1	83	85
Manual Omissions	≥25.5	25					100	59
(% of non-DCR trials)	20.8	29					100	60
	17.4	29	4	64	7.6	18.4	88	60
	17.3	33	4	66	8.7	17.3	89	61
	17.1	38	4	68	9.8	16.3	90	63
	14.1	42	4	70	10.8	15.2	91	64
	10.7	46	4	72	11.9	14.1	92	66
	9.7	50	4	74	13.0	13.0	92	68
	8.2	54	4	76	14.1	11.9	93	69
	7.1	58	4	78	15.2	10.8	93	71
	5.9	67	4	82	17.3	8.7	94	76
	5.2	71	4	84	18.4	7.6	94	78
	4.3	75	4	86	19.5	6.5	95	81
	3.3	79	4	88	20.6	5.4	95	83
	3.0	83	4	90	21.7	4.3	95	86
	2.7	88	4	92	22.8	3.3	95	89
	2.4	92	4	94	23.8	2.2	96	93
	1.7	92	8	92	11.9	1.1	92	92
	1.0	92	12	90	7.9	0.7	88	92

Note. BR = biased responders; UR = unbiased responders; LR = likelihood ratio; Sac = saccadic; Man = manual; RT = reaction time; IIV = intra-individual variability; DC = directional cue; NDC = nondirectional cue; MDC = misdirectional cue; UCG = uncued with gap; UC = uncued; DCR = directional cue-red arrow.

[a]n=24 for all BR variables except Saccadic RT-IIV-UCG (n=23), Manual RT-IIV-DC (n=23), Manual RT-IIV-NDC (n=23), Manual RT-IIV-MDC (n=18), Manual RT-IIV-UCG (n=21), Manual RT-IIV-UC (n=23), and Manual RT-IIV-Overall (n=23).

[b]n=26 for all UR variables.

Table 8: ROC analyses for embedded RVT variables in the simulator study

Variable	Positive (BR)	Negative (UR)	AUC	SE	95% Low	95% Hi	*p*
WAIS-IV Digit Span RDS-R	24	26	0.94	0.03	0.88	1.00	<.001
WAIS-IV Digit Span ACSS	24	26	0.94	0.03	0.88	1.00	<.001
WAIS-IV Digit Span ARDS	24	26	0.94	0.04	0.86	1.00	<.001
WAIS-IV Digit Span RDS	24	26	0.93	0.04	0.84	1.00	<.001
CPT-II Commissions (raw)	24	26	0.93	0.04	0.85	1.00	<.001
CPT-II Omissions (raw)	24	26	0.91	0.05	0.82	1.00	<.001
CPT-II Hit RT SE (raw)	24	26	0.90	0.04	0.82	0.99	<.001
CPT-II Variability (raw)	24	26	0.89	0.05	0.80	0.98	<.001
CPT-II Detectability (raw)	24	26	0.88	0.05	0.77	0.98	<.001
CPT-II Perseverations (raw)	24	26	0.84	0.06	0.73	0.96	<.001
CPT-II RT ISI Change (raw)	24	26	0.79	0.06	0.67	0.91	<.001
Trail Making Test B Time (sec)	24	26	0.79	0.06	0.66	0.91	.001
Trail Making Test A Time (sec)	24	26	0.74	0.08	0.59	0.90	.003
CPT-II SE ISI Change (raw)	24	26	0.67	0.08	0.52	0.82	.04
CPT-II Response Style (raw)	24	26	0.67	0.08	0.51	0.82	.05
CPT-II Hit RT SE Block Change (raw)	24	26	0.54	0.08	0.37	0.70	.67
CPT-II Hit RT (raw)	24	26	0.52	0.09	0.35	0.69	.83
CPT-II Hit RT Block Change (raw)	24	26	0.50	0.09	0.33	0.66	.96

Note. WAIS-IV = Wechsler Adult Intelligence Scale-4th Edition; RDS-R = Reliable Digit Span-Revised; ACSS = Age-Corrected Scaled Score; ARDS = Alternative Reliable Digit Span; RDS = Reliable Digit Span; CPT-II = Conners' Continuous Performance Test-2nd Edition; RT = reaction time; SE = standard error; ISI = interstimulus interval.

Table 9: Non-normally distributed embedded RVT variables with outstanding classification accuracy

	Unbiased responders (n = 26)		Biased responders (n = 24)			
Measures	*Mdn*	*IQR*	*Mdn*	*IQR*	*p*	*r*
WAIS-IV Digit Span						
RDS-R	17.0	16.0 - 18.0	11.0	10.0 - 13.8	<.001	.77
ARDS	12.0	11.8 - 13.0	8.50	7.00 - 10.0	<.001	.76
CPT-II						
Omissions (raw)	0.00	0.00 - 1.00	6.50	2.00 - 11.0	<.001	.72
Hit RT SE (raw)	4.20	3.30 - 4.82	7.56	5.33 - 9.91	<.001	.69

Note. Mann-Whitney *U* effect size "r" was calculated by dividing Z-score by the square root of *N*. WAIS-IV = Wechsler Adult Intelligence Scale-4th Edition; RDS-R = Reliable Digit Span-Revised; ARDS = Alternative Reliable Digit Span; RT = reaction time; SE = standard error.

Table 10: Normally distributed embedded RVT variables with outstanding classification accuracy

	Unbiased responders (n = 26)		Biased responders (n = 24)			
Measures	*M*	*SD*	*M*	*SD*	*p*	Cohen's *d*
WAIS-IV Digit Span						
ACSS	12.2	2.43	6.63	2.57	<.001	2.23
RDS	11.1	1.80	7.29	2.12	<.001	1.94
Conners' CPT-II						
Commissions (raw)	10.4	4.00	20.6	5.34	<.001	2.16

Note. WAIS-IV = Wechsler Adult Intelligence Scale-4th Edition; ACSS = Age-Corrected Scaled Score; RDS = Reliable Digit Span.

Table 11: Cumulative percentages of persons with scores above or below the indicated cutoff on selected embedded RVT variables

Test	*Variable*	*Cutoff*	*% BR (n=24)*	*% UR (n=26)*	*Hit Rate*	*LR+*	*LR-*	*PPV*	*NPV*
WAIS-IV:	ACSS	≤4	29					100	60
Digit Span		5	33					100	62
		6	46	4	72	11.9	14.1	92	66
		7	67	4	82	17.3	8.7	94	76
		8	71	4	84	18.4	7.6	94	78
		9	92	12	90	7.9	0.7	88	92
		10	92	15	88	6.0	0.5	85	92
	RDS	≤5	25					100	59
		6	42					100	65
		7	54	4	76	14.1	11.9	93	69
		8	67	4	82	17.3	8.7	94	76
		9	92	12	90	7.9	0.7	88	92
	RDS-R	≤9	21					100	58
		10	33					100	62
		11	54					100	70
		12	63	4	80	16.3	9.8	94	74
		13	75	4	86	19.5	6.5	95	81
		14	88	8	90	11.4	1.6	91	89
		15	92	12	90	7.9	0.7	88	92
	ARDS	≤7	33					100	62
		8	50					100	68
		9	67					100	76
		10	88	8	90	11.4	1.6	91	89
CPT-II	Commissions (raw)	≥24	29					100	60
		22	42					100	65
		21	67					100	76
		20	71					100	79
		19	75	4	86	19.5	6.5	95	81
		18	79	4	88	20.6	5.4	95	83
		17	79	12	84	6.9	1.8	86	82
		16	83	12	86	7.2	1.4	87	85
		15	83	15	84	5.4	1.1	83	85
	Omissions (raw)	≥9	42					100	65
		7	50					100	68
		6	63					100	74
		5	67					100	76
		4	67	4	82	17.3	8.7	94	76
		3	71	15	78	4.6	1.9	81	76
		2	92	15	88	6.0	0.5	85	92
	Hit RT SE (sec)	≥9.2	38					100	63
		8.8	42					100	65
		8.4	46					100	67
		7.6	50					100	68
		6.7	50	4	74	13.2	13.2	92	68
		5.8	71	4	84	18.6	7.7	95	78
		5.6	71	8	82	9.2	3.8	89	77
		5.4	75	8	84	9.7	3.2	90	80
		5.3	75	12	82	6.5	2.2	86	79
		5.2	79	12	84	6.9	1.8	86	82
		5.1	79	15	82	5.1	1.4	83	82
		5.0	88	15	86	5.7	0.8	84	88

Note. BR = biased responders; UR = unbiased responders; LR = likelihood ratio; WAIS-IV = Wechsler Adult Intelligence Scale-4th Edition; ACSS = age-corrected scaled score; RDS = reliable digit span; RDS-R = reliable digit span-revised; ARDS = alternative reliable digit span; CPT-II = Conners' Continuous Performance Test-II; RT = reaction time; SE = standard error.

Table 12: ROC analyses for freestanding RVT variables

Variable	Positive (BR)	Negative (UR)	AUC	SE	95% Low	95% Hi	*p*
VSVT Difficult Correct	24	26	1.00	0.00	1.00	1.00	<.001
VSVT Total Correct	24	26	1.00	0.00	1.00	1.00	<.001
MSVT FR	24	26	0.96	0.03	0.91	1.00	<.001
MSVT CNS	24	26	0.94	0.04	0.87	1.00	<.001
MSVT PA	24	26	0.93	0.04	0.85	1.00	<.001
VSVT Total RT (sec)	24	26	0.93	0.04	0.86	1.00	<.001
MSVT DR	24	26	0.92	0.04	0.84	1.00	<.001
VSVT Difficult RT (sec)	24	26	0.91	0.05	0.82	0.99	<.001
VSVT Easy RT (sec)	24	26	0.90	0.04	0.81	0.98	<.001
VSVT Difficult RT SD (sec)	24	26	0.90	0.05	0.81	0.99	<.001
MSVT IR	24	26	0.90	0.05	0.80	1.00	<.001
MSVT Fail Any Subtest (per norms)	24	26	0.90	0.05	0.80	1.00	<.001
VSVT Total RT SD (sec)	24	26	0.87	0.05	0.77	0.98	<.001
VSVT Easy RT SD (sec)	24	26	0.85	0.06	0.74	0.96	<.001
VSVT Easy Correct	24	26	0.85	0.06	0.73	0.96	<.001

Note: VSVT = Victoria Symptom Validity Test; MSVT = Medical Symptom Validity Test; FR = Free Recall % Correct; CNS = Consistency % Correct; PA = Paired Associates % Correct; RT = response latency; DR = Delayed Recall % Correct; SD = standard deviation; IR = Immediate Recall % Correct.

Table 13: Non-normally distributed freestanding RVT variables with outstanding classification accuracy

	Unbiased responders (n = 26)		Biased responders (n = 24)			
Measures	*Mdn*	*IQR*	*Mdn*	*IQR*	*p*	*r*
VSVT						
Difficult Correct	24.0	22.8 - 24.0	12.5	10.0 - 13.0	<.001	.87
Total Correct	48.0	46.8 - 48.0	32.5	27.8 - 36.8	<.001	.87
Easy RT (sec)	1.02	0.88 - 1.10	1.93	1.31 - 2.96	<.001	.68
Difficult RT (sec)	1.72	1.43 - 1.90	3.45	2.26 - 5.36	<.001	.69
Difficult RT SD (sec)	0.49	0.38 - 0.64	1.14	0.76 - 2.72	<.001	.68
Total RT (sec)	1.38	1.14 - 1.51	2.91	1.68 - 4.38	<.001	.73
MSVT						
IR	100	100 - 100	85.0	75.0-95.0	<.001	.78
DR	100	100 - 100	75.0	56.3-87.8	<.001	.78
CNS	100	100 - 100	65.0	60.0 - 93.8	<.001	.81
PA	100	100 - 100	70.0	60.0 - 87.5	<.001	.82
FR	90.0	80.0 - 90.0	55.0	41.3 - 63.8	<.001	.79
Fail Any Subtest	0.00	0.00 - 0.00	1.00	1.30 - 4.00	<.001	.81[a]

Note: VSVT = Victoria Symptom Validity Test; RT = response latency; SD = standard deviation; IR; MSVT = Medical Symptom Validity Test; FR = Free Recall % Correct; CNS = Consistency % Correct; PA = Paired Associates % Correct; DR = Delayed Recall % Correct; = Immediate Recall % Correct.
[a]Phi value

Table 14: Cumulative percentages of persons with scores above or below the indicated cutoff on selected freestanding RVT variables

Test	*Variable*	*Cutoff*	*% BR (n=24)*	*% UR (n=26)*	*Hit Rate*	*LR+*	*LR-*	*PPV*	*NPV*
VSVT	Easy RT (sec)	≥1.99	50	0	76			100	68
		1.93	50	4	74	13.0	13.0	92	68
		1.87	50	8	72	6.5	6.5	86	67
		1.78	54	8	74	7.0	6.0	87	69
		1.70	58	8	76	7.6	5.4	88	71
		1.59	63	8	78	8.1	4.9	88	73
		1.51	63	12	76	5.4	3.3	83	72
		1.49	67	12	78	5.8	2.9	84	74
		1.40	71	12	80	6.1	2.5	85	77
		1.31	75	12	82	6.5	2.2	86	79
		1.25	79	12	84	6.9	1.8	86	82
		1.17	83	12	86	7.2	1.4	87	85
		1.14	88	12	88	7.6	1.1	88	88
	Difficult RT (sec)	≥3.28	54	0	78			100	70
		3.09	54	4	76	14.1	11.9	93	69
		2.99	58	4	78	15.2	10.8	93	71
		2.78	63	4	80	16.3	9.8	94	74
		2.53	67	4	82	17.3	8.7	94	76
		2.36	67	8	80	8.7	4.3	89	75
		2.29	75	8	84	9.8	3.3	90	80
		2.26	75	12	82	6.5	2.2	86	79
		2.15	79	12	84	6.9	1.8	86	82
		2.02	83	12	86	7.2	1.4	87	85
	Difficult RT SD (sec)	≥1.57	46		74			100	67
		1.43	46	4	72	12.1	14.3	92	66
		1.15	50	4	74	13.2	13.2	92	68
		0.94	50	8	72	6.5	6.5	86	67
		0.93	54	8	74	7.0	5.9	87	69
		0.91	58	8	76	7.6	5.4	87	71
		0.89	58	12	74	5.1	3.6	82	70
		0.86	67	12	78	5.8	2.9	84	74
		0.80	71	12	80	6.2	2.5	85	77
		0.76	75	12	82	6.5	2.2	86	79
		0.73	79	12	84	6.9	1.8	86	82
	Total RT (sec)	≥2.67	58	0	80			100	72
		2.38	58	4	78	15.2	10.8	93	71
		2.14	63	4	80	16.3	9.8	94	74
		1.97	67	4	82	17.3	8.7	94	76
		1.83	71	4	84	18.4	7.6	94	78
		1.74	71	8	82	9.2	3.8	89	77
		1.70	75	8	84	9.8	3.3	90	80
		1.68	75	12	82	6.5	2.2	86	79
		1.66	79	12	84	6.9	1.8	86	82
		1.62	88	12	88	7.6	1.1	88	88
		1.60	88	15	86	5.7	0.8	84	88
	Difficult Correct	≤12	50	0				100	68
		13	79	0				100	84
		14	88	0				100	90
		15	92	0				100	93
		18	100	0				100	100
		21	100	15	92	6.5	0	86	100
	Total Correct	≤35	67	0				100	76
		36	75	0				100	81
		38	92	0				100	93
		39	96	0				100	96
		42	100	0				100	100
		45	100	15	92	6.5	0	86	100

Test	*Variable*	*Cutoff*	*% BR (n=24)*	*% UR (n=26)*	*Hit Rate*	*LR+*	*LR-*	*PPV*	*NPV*
MSVT	IR (% correct)	≤70	13	0	58			100	55
		75	38	0	70			100	63
		80	42	0	72			100	65
		85	54	0	78			100	70
		90	71	0	86			100	79
		95	79	0	90			100	84
	DR (% correct)	≤65	38	0	70			100	63
		70	46	0	74			100	67
		75	54	0	78			100	70
		85	75	0	88			100	81
		90	79	4	88	20.6	5.4	95	83
		95	88	12	88	7.6	1.1	88	88
	CNS (% correct)	≤70	54	0	78			100	70
		75	63	0	82			100	74
		80	67	0	84			100	76
		85	71	0	86			100	79
		90	75	4	86	19.5	6.5	95	81
		95	92	12	90	7.9	0.7	88	92
	PA (% correct)	≤45	8	0	56			100	54
		55	17	0	60			100	57
		65	38	0	70			100	63
		75	67	0	84			100	76
		85	75	0	88			100	81
		95	88	4	92	22.75	3.25	95	89
	FR (% correct)	≤45	29	0	66			100	60
		50	42	0	72			100	65
		55	54	0	78			100	70
		60	75	4	86	19.5	6.5	95	81
		65	83	4	90	21. 7	4.3	95	86
	Fail Any Subtest	5	0	0	52			100	52
	(# subtests failed)	4	13	0	58			100	55
		3	54	0	78			100	70
		2	71	0	86			100	79
		1	75	0	88			100	81

Note: VSVT = Victoria Symptom Validity Test; RT = response latency; SD = standard deviation; IR; MSVT = Medical Symptom Validity Test; FR = Free Recall % Correct; CNS = Consistency % Correct; PA = Paired Associates % Correct; DR = Delayed Recall % Correct; = Immediate Recall % Correct.

Table 15: Hierarchical logistic regressions examining incremental validity of representative variables

Model	Block	Scale	Model χ^2 (*df*)	*p*	χ^2-change (*df*)	*p*	R^2	R^2-change
Embedded→ BEAM	1	RDS-R	40.7 (1)	<.001			.74	
	2	SacRT-IIV-Overall	58.1 (2)	<.001	17.4 (1)	<.001	.92	.18
Freestanding[a]→ BEAM	1	MSVT FR	39.8 (1)	<.001			.74	
	2	ManRT-IIV-Overall	53.3 (2)	<.001	13.5 (1)	<.001	.89	.15
Embedded→ Freestanding[a]→ BEAM	1	RDS-R	40.7 (1)	<.001			.74	
	2	MSVT FR	52.5 (2)	<.001	11.8 (1)	.001	.87	.13
	3	SacRT-IIV-Overall	*b*	*b*	*b*	*b*	*b*	*b*

Note: RDS-R = Reliable Digit Span-Revised; SacRT-IIV = saccadic reaction time intra-individual variability; MSVT = Medical Symptom Validity Test; FR = Free Recall % Correct; ManRT-IIV = manual reaction time intra-individual variability.
[a]VSVT Difficult Correct and VSVT Total Correct had better classification accuracy than MSVT FR but could not be loaded into the regression models.
[b]The model achieved exact classification accuracy and values could not be calculated.

Table 16: Demographic characteristics of clinical comparison group

	UR-mTBI (n=19)	***F*[a]**	***p***
Mean age in years (SD)	34.3 (13.3)	1.96	.15
Mean years of education (SD)	16.2 (2.4)	0.67	.52
Estimated premorbid intelligence (SD)	114 (7.7)	0.77	.47
		χ^2(df)[a]	
Gender (%)		0.54 (2)	.76
Male	7 (36.8)		
Female	12 (63.2)		
Race/ethnicity (%)		9.22 (8)	.32
Caucasian	15 (78.9)		
African-American	2 (10.5)		
Hispanic	1 (5.3)		
Asian	1 (5.3)		
Injury characteristics			
Median years since injury (IQR)	6.9 (2.32 - 21.6)		
LOC length in minutes (SD)	3.00 (4.29)		
PTA length in minutes (SD)	18.1 (50.8)		

Note: UR-mTBI = unbiased responders with a history of mild TBI; IQR = interquartile range; LOC = loss of consciousness; PTA = posttraumatic amnesia; SD = standard deviation.
[a]One-way ANOVA or chi-square with BR, UR, and UR-mTBI groups.

Table 17: Group comparisons of normally distributed BEAM and embedded RVT variables with excellent classification accuracy

		1) UR			2) UR-mTBI			3) BR			ANOVA		Cohen's effect sizes (*d*)		
Test	**Measures**	**n**	***M***	***SD***	**n**	***M***	***SD***	**n**	***M***	***SD***	***p***	***Pairwise comparisons***	***UR vs UR-mTBI***	***UR vs BR***	***UR-mTBI vs BR***
BEAM	SacRT-IIV-DC (sec)	26	0.078	0.024	19	0.088	0.023	24	0.139	0.031	<.001	1 & 2 < 3	.43	2.20	1.87
	SacRT-IIV-MDC (sec)	26	0.080	0.023	19	0.101	0.032	24	0.160	0.050	<.001	1 & 2 < 3	*a*	*a*	*a*
	SacRT-IIV-UCG (sec)	26	0.076	0.021	19	0.088	0.031	23	0.137	0.038	<.001	1 & 2 < 3	.45	1.99	1.41
	SacRT-IIV-Overall (sec)	26	0.082	0.018	19	0.094	0.024	24	0.142	0.026	<.001	1 & 2 < 3	.56	2.68	1.92
	ManRT-IIV-DC (sec)	26	0.076	0.021	19	0.087	0.028	23	0.151	0.030	<.001	1 & 2 < 3	.44	2.90	2.21
	ManRT-IIV-NDC (sec)	26	0.065	0.022	19	0.079	0.027	23	0.128	0.027	<.001	1 & 2 < 3	.57	2.56	1.81
	ManRT-IIV-MDC (sec)	26	0.077	0.025	19	0.084	0.026	18	0.144	0.034	<.001	1 & 2 < 3	.27	2.25	1.98
	ManRT-IIV-Overall (sec)	26	0.076	0.018	19	0.086	0.023	23	0.134	0.022	<.001	1 & 2 < 3	.48	2.89	2.13
WAIS-IV	ACSS	26	12.2	2.43	19	11.4	2.00	24	6.63	2.57	<.001	1 & 2 < 3	.36	2.23	2.07
	RDS	26	11.1	1.80	19	11.1	1.65	24	7.29	2.12	<.001	1 & 2 < 3	.04	1.94	2.02
CPT-II	Commissions (raw)	26	10.4	4.00	19	13.6	4.97	24	20.6	5.34	<.001	1 & 2 < 3	.71	2.16	1.36

Note. UR = unbiased responders; UR-mTBI = unbiased responders with a history of mild TBI; BR = biased responders; BEAM = Bethesda Eye & Attention Measure; Sac = saccadic; RT = reaction time; IIV = intra-individual variability; DC = directional cue; MDC = misdirectional cue; UCG = uncued with gap; Man = manual; NDC = nondirectional cue; WAIS-IV = Wechsler Adult Intelligence Scale-4th Edition; ACSS = Age Corrected Scaled Score; RDS = Reliable Digit Span; CPT-II = Conners' Continuous Performance test-2nd Edition.
[a]Effect size could not be calculated since homogeneity of variance assumption was violated

Table 18: Group comparisons of non-normally distributed BEAM and embedded RVT variables with excellent classification accuracy

		1) UR			2) UR-mTBI			3) BR			Kruskal-Wallis		*r*		
Test	**Measures**	**n**	***Mdn***	***IQR***	**n**	***Mdn***	***IQR***	**n**	***Mdn***	***IQR***	***p***	***Pairwise comparisons***	***UR vs UR-mTBI***	***UR vs BR***	***UR-mTBI vs BR***
BEAM	SacCom%	26	7.29	2.42 - 17.6	19	8.33	0.00 - 32.0	24	65.5	45.8 - 79.9	<.001	1 & 2 < 3	.03	.75	.69
	ManOm%	26	0.00	0.00 - 0.63	19	0.00	0.00 - 0.63	24	9.65	3.91 - 26.1	<.001	1 & 2 < 3	.05	.78	.71
	ManRT-IIV-UCG	26	0.072	0.059 - 0.086	19	0.080	0.068 - 0.115	21	0.128	0.112 - 0.128	<.001	1 & 2 < 3	.23	.72	.55
	ManRT-IIV-UC	26	0.085	0.075 - 0.095	19	0.098	0.077 - 0.107	23	0.130	0.103 - 0.150	<.001	1 & 2 < 3	.20	.71	.54
WAIS-IV	RDS-R	26	17.0	16.0 - 18.0	19	16.0	15.0 - 18.0	24	11.0	10.0 - 13.8	<.001	1 & 2 < 3	.26	.77	.71
	ARDS	26	12.0	11.8 - 13.0	19	11.0	10.0 - 13.0	24	8.50	7.00 - 10.0	<.001	1 & 2 < 3	.25	.76	.62
CPT-II	Omissions (raw)	26	0.00	0.00 - 1.00	19	1.00	0.00 - 2.00	24	6.50	2.00 - 11.0	<.001	1 & 2 < 3	.23	.72	.65
	Hit RT SE (raw)	26	4.20	3.30 - 4.82	19	4.20	3.68 - 5.32	24	7.56	5.33 - 9.91	<.001	1 & 2 < 3	.20	.69	.60

Note. UR = unbiased responders; UR-mTBI = unbiased responders with a history of mild TBI; BR = biased responders; BEAM = Bethesda Eye & Attention Measure; Sac = saccadic; RT = reaction time; IIV = intra-individual variability; DC = directional cue; MDC = misdirectional cue; UCG = uncued with gap; Man = manual; NDC = nondirectional cue; WAIS-IV = Wechsler Adult Intelligence Scale-4th Edition; ACSS = Age Corrected Scaled Score; RDS = Reliable Digit Span; CPT-ii = Conners' Continuous Performance test-2nd Edition.

[a]Effect size could not be calculated since homogeneity of variance assumption was violated

Table 19: Cumulative percentages of persons with scores above the indicated cutoff on selected BEAM variables

Variable	*Cutoff*	*BR[a] (%)*	*UR-mTBI[b] (%)*	*UR[c] (%)*
SacRT-IIV-DC (sec)	≥0.154	38		
	0.153	38	5	
	0.151	42	5	
	0.147	46	5	
	0.142	50	5	
	0.137	54	5	
	0.131	58	5	
	0.130	63	5	
	0.129	67	5	
	0.128	71	5	
	0.127	71	5	4
	0.126	75	5	4
	0.125	79	5	4
	0.119	79	5	8
	0.113	79	5	12
	0.110	79	11	12
	0.109	79	11	15
	0.104	79	16	15
SacRT-IIV-MDC (sec)	≥0.189	29		
	0.179	29	5	
	0.173	33	5	
	0.169	38	5	
	0.167	42	5	
	0.165	46	5	
	0.162	50	5	
	0.159	54	5	
	0.152	58	5	
	0.144	63	5	
	0.142	67	5	
	0.139	71	5	
	0.135	71	11	
	0.134	75	11	
	0.133	75	16	
	0.131	75	21	
	0.129	79	21	
	0.126	83	26	
	0.123	83	26	4
	0.122	88	26	4
	0.119	88	26	8
	0.115	88	26	12
	0.113	88	32	12
	0.112	88	32	15
SacRT-IIV-UCG (sec)	≥0.177	13		
	0.174	13	5	
	0.169	17	5	
	0.162	22	5	
	0.157	26	5	
	0.154	30	5	
	0.152	35	5	
	0.151	39	5	
	0.148	43	5	
	0.146	48	5	
	0.143	52	5	
	0.138	57	5	
	0.137	61	5	
	0.136	65	5	
	0.131	70	5	
	0.125	74	11	

Variable	*Cutoff*	*BR[a] (%)*	*UR-mTBI[b] (%)*	*UR[c] (%)*
	0.122	78	11	
	0.117	78	11	4
	0.115	78	16	4
	0.113	78	21	4
	0.109	78	26	8
	0.102	78	32	12
	0.099	78	32	15
	0.097	83	32	15
	0.095	87	32	15
SacRT-IIV-Overall (sec)	≥0.135	63		
	0.132	67	5	
	0.130	71	11	
	0.125	75	16	
	0.120	79	21	
	0.118	79	21	
	0.117	83	21	4
	0.116	83	21	4
	0.115	88	21	8
	0.113	88	26	8
	0.112	88	26	12
	0.108	88	37	15
	0.106	92	37	15
	0.102	96	42	15
ManRT-IIV-DC (sec)	≥0.137	74		
	0.130	78		
	0.127	83		
	0.126	83		
	0.123	83	5	4
	0.122	83	11	4
	0.121	83	16	4
	0.120	83	21	8
	0.117	87	26	8
	0.112	91	32	8
	0.105	91	32	8
	0.094	91	37	12
	0.093	91	37	15
ManRT-IIV-NDC (sec)	≥0.141	39	0	
	0.140	39	5	
	0.138	43	5	
	0.134	48	5	
	0.129	52	5	
	0.126	57	5	
	0.119	61	5	
	0.117	61	11	
	0.113	65	11	
	0.111	65	11	4
	0.110	70	11	4
	0.109	74	11	4
	0.108	78	11	4
	0.106	83	11	4
	0.104	87	11	4
	0.103	87	16	4
	0.101	91	16	4
	0.099	91	21	8
	0.094	91	26	15
	0.092	96	26	15
ManRT-IIV-MDC (sec)	≥0.134	56		
	0.131	56	5	
	0.127	61	5	
	0.125	67	5	
	0.123	72	5	

Variable	*Cutoff*	*BR[a] (%)*	*UR-mTBI[b] (%)*	*UR[c] (%)*
	0.120	72	11	4
	0.118	72	11	8
	0.117	72	16	8
	0.116	78	16	8
	0.114	78	16	12
	0.111	83	16	15
ManRT-IIV-UCG (sec)	≥0.151	5		
	0.148	5	5	
	0.147	10	5	
	0.145	14	5	
	0.142	14	11	
	0.140	19	11	
	0.138	24	11	
	0.135	29	11	
	0.132	33	11	
	0.131	38	11	
	0.130	43	11	
	0.129	43	11	4
	0.128	48	11	4
	0.127	52	11	4
	0.126	57	11	4
	0.125	62	11	4
	0.123	67	16	4
	0.121	67	21	4
	0.118	71	21	4
	0.115	76	21	4
	0.112	76	26	8
	0.107	81	26	8
	0.103	86	26	8
	0.100	86	26	12
	0.097	86	32	15
	0.096	90	32	15
	0.095	95	32	15
ManRT-IIV-UC (sec)	≥0.150	22		
	0.146	26		
	0.140	26		4
	0.139	26	5	4
	0.136	30	5	4
	0.134	35	5	4
	0.132	39	5	4
	0.131	43	5	4
	0.130	48	5	4
	0.126	52	5	4
	0.122	57	5	4
	0.121	57	11	4
	0.119	61	11	4
	0.117	61	11	8
	0.116	61	16	8
	0.111	65	16	8
	0.108	65	21	8
	0.107	65	26	8
	0.105	65	32	12
	0.104	74	32	15
ManRT-IIV-Overall (sec)	≥0.128	70		
	0.127	74		
	0.122	78		
	0.118	78	5	4
	0.117	78	11	8
	0.113	83	16	8
	0.111	83	16	8
	0.110	83	21	8

Variable	*Cutoff*	*BR*[a] *(%)*	*UR-mTBI*[b] *(%)*	*UR*[c] *(%)*
	0.106	87	21	8
	0.104	91	26	8
	0.100	91	32	8
	0.099	91	32	8
	0.094	96	37	8
	0.093	96	37	12
	0.092	96	37	15
Saccadic Commissions	≥61.5	54		
(% of DCR trials)	57.9	58	5	
	55.4	63	5	
	53.3	67	5	
	52.6	71	5	
	47.9	75	5	
	45.7	75	11	
	43.5	79	16	
	43.1	83	16	4
	41.8	83	16	8
	33.2	83	21	8
	32.9	83	21	12
	24.0	83	26	15
Manual Omissions	≥30.4	21		
(% of non-DCR trials)	28.5	21	5	
	25.5	25	5	
	20.8	29	5	
	17.4	29	5	4
	17.3	33	5	4
	17.1	38	5	4
	14.1	42	5	4
	10.7	46	5	4
	9.7	50	5	4
	8.2	54	5	4
	7.9	54	11	4
	7.1	58	11	4
	5.9	67	11	4
	5.2	71	11	4
	4.3	75	11	4
	3.3	79	11	4
	3.0	83	11	4
	2.7	88	11	4
	2.4	92	11	4
	1.7	92	11	8
	1.0	92	11	12

Note. BR = biased responders; UR = unbiased responders; Sac = saccadic; Man = manual; RT = reaction time; IIV = intra-individual variability; DC = directional cue; NDC = nondirectional cue; MDC = misdirectional cue; UCG = uncued with gap; UC = uncued.
[a]n=24 for all BR variables except Saccadic RT-IIV-UCG (n=23), Manual RT-IIV-DC (n=23), Manual RT-IIV-NDC (n=23), Manual RT-IIV-MDC (n=18), Manual RT-IIV-UCG (n=21), Manual RT-IIV-UC (n=23), and Manual RT-IIV-Overall (n=23).
[b]n=19 for all UR-mTBI variables.
[b]n=26 for all UR variables.

Table 20: Cumulative percentages of persons with scores above or below the indicated cutoff on selected embedded RVT variables

Test	*Variable*	*Cutoff*	*BR (n=24)*	*UR-mTBI (n=19)*	*UR (n=26)*
WAIS-IV:	ACSS	≤5	33		
Digit Span		6	46		4
		7	67		4
		8	71	5	4
		9	92	16	12
		10	92	32	15
	RDS	≤6	42		
		7	54		4
		8	67	5	4
		9	92	16	12
	RDS-R	≤11	54		
		12	63		4
		13	75	5	4
		14	88	16	8
		15	92	37	12
	ARDS	≤7	33		
		8	50	5	
		9	67	16	
		10	88	26	8
		11	96	53	23
CPT-II	Commissions (raw)	≥22	42		
		21	67	5	
		20	71	16	
		19	75	16	4
		18	79	26	4
		17	79	32	12
		16	83	42	12
		15	83	47	15
	Omissions (raw)	≥6	63		
		5	67	5	
		4	67	11	4
		3	71	16	15
	Hit RT SE (sec)	≥7.9	50		
		7.2	50	5	
		6.9	50	11	
		6.7	50	11	4
		6.4	58	11	4
		6.0	58	16	4
		5.6	71	16	8
		5.5	71	21	8
		5.3	75	26	12
		5.0	88	26	15

Note. BR = biased responders; UR=mTBI = unbiased responders with a history of mild TBI; UR = unbiased responders; WAIS-IV = Wechsler Adult Intelligence Scale-4th Edition; ACSS = Age Corrected Scaled Score; RDS = Reliable Digit Span; RDS-R = Reliable Digit Span-Revised; ARDS = Alternative Reliable Digit Span; CPT-II = Conners' Continuous Performance Test-II.

APPENDIX B: FIGURES

Figure 1: Stepwise Logistic Regression Illustration for BEAM Variables

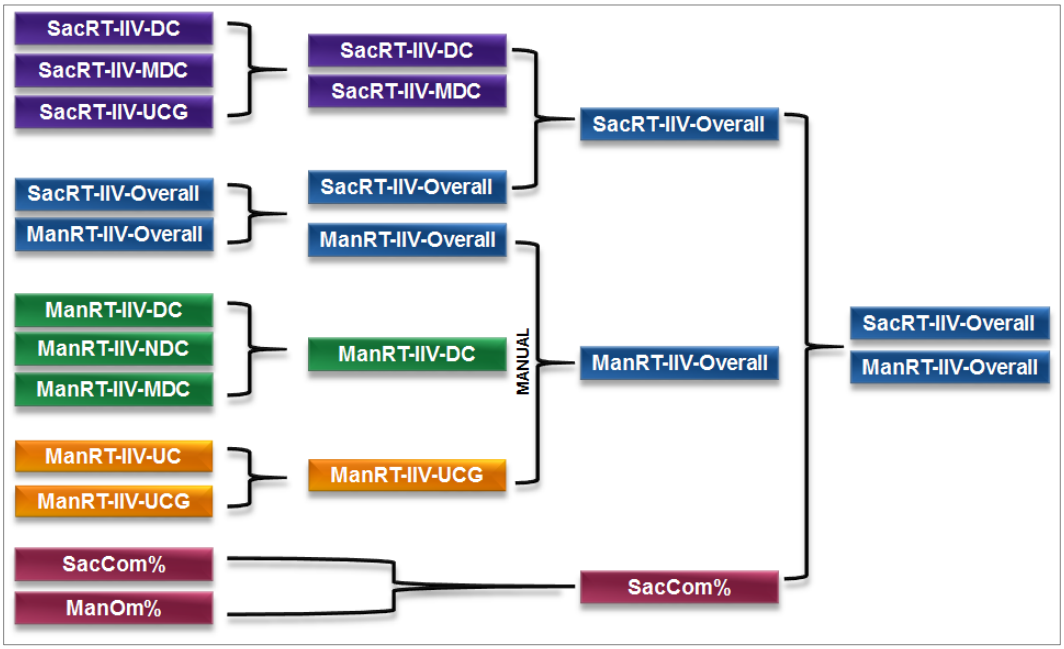


Figure 2: Variable Reduction Depiction for Simulator Study Variables

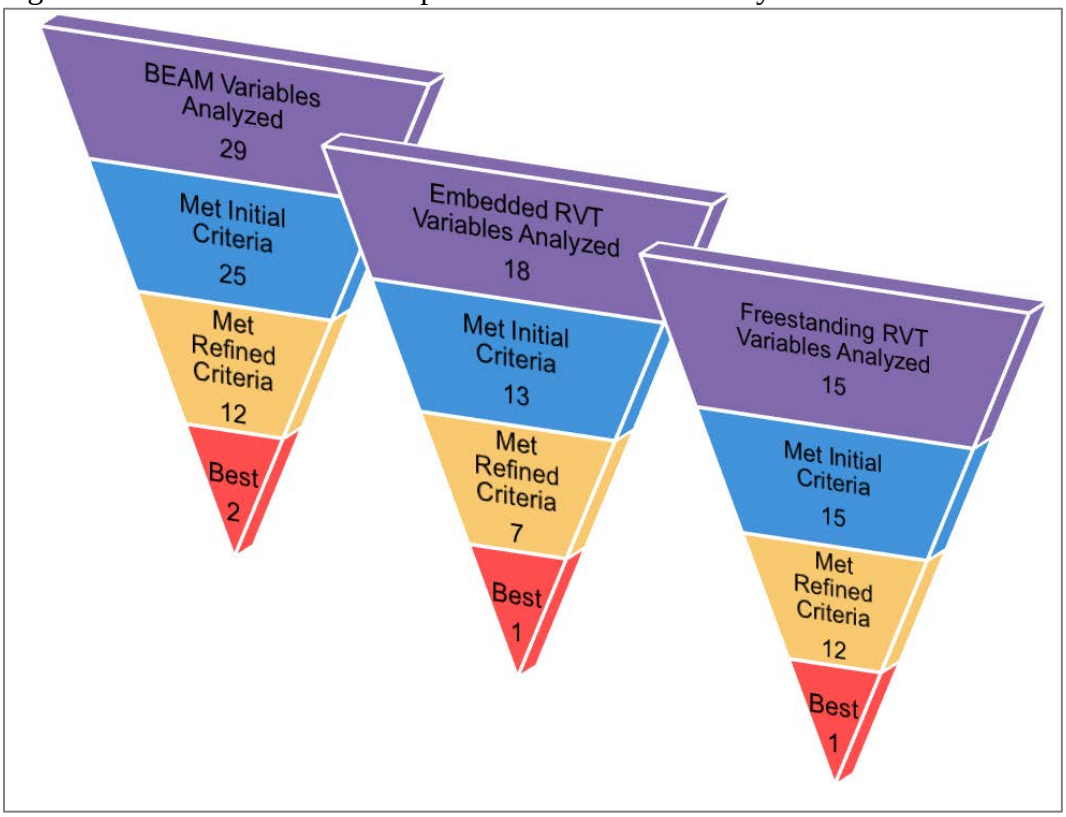


Figure 3: Saccadic RT-IIV-Directional Cue Sensitivity and Specificity by Cutoff Score

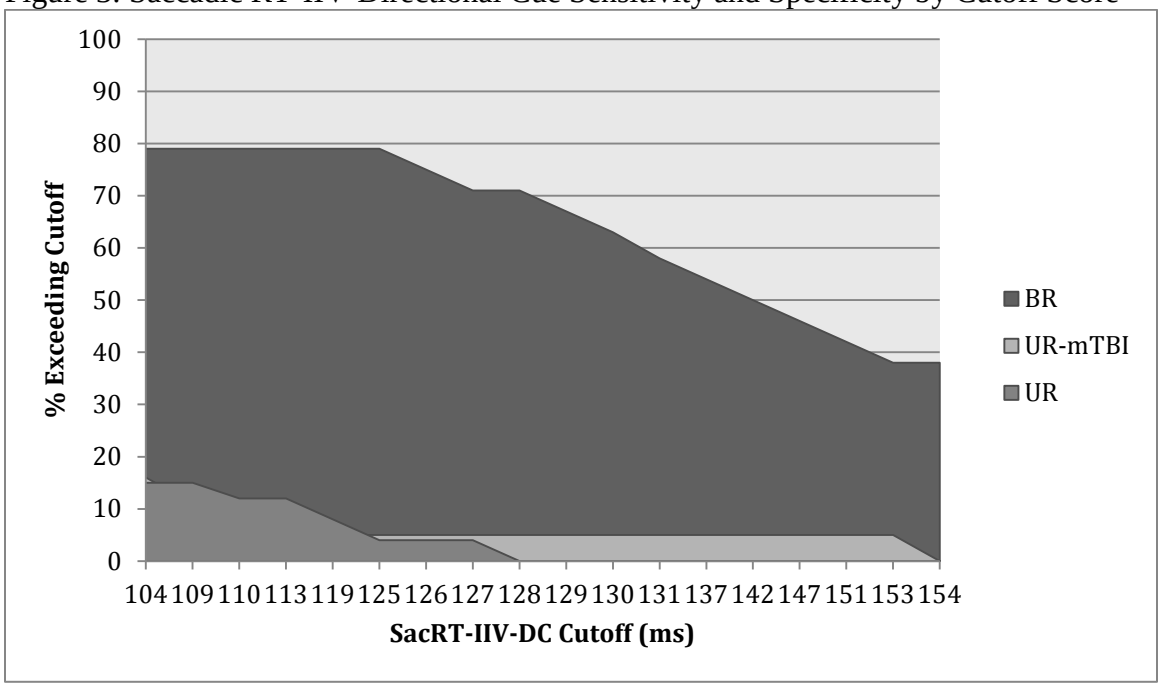


Figure 4: Manual Omission Error Rate Sensitivity and Specificity by Cutoff Score

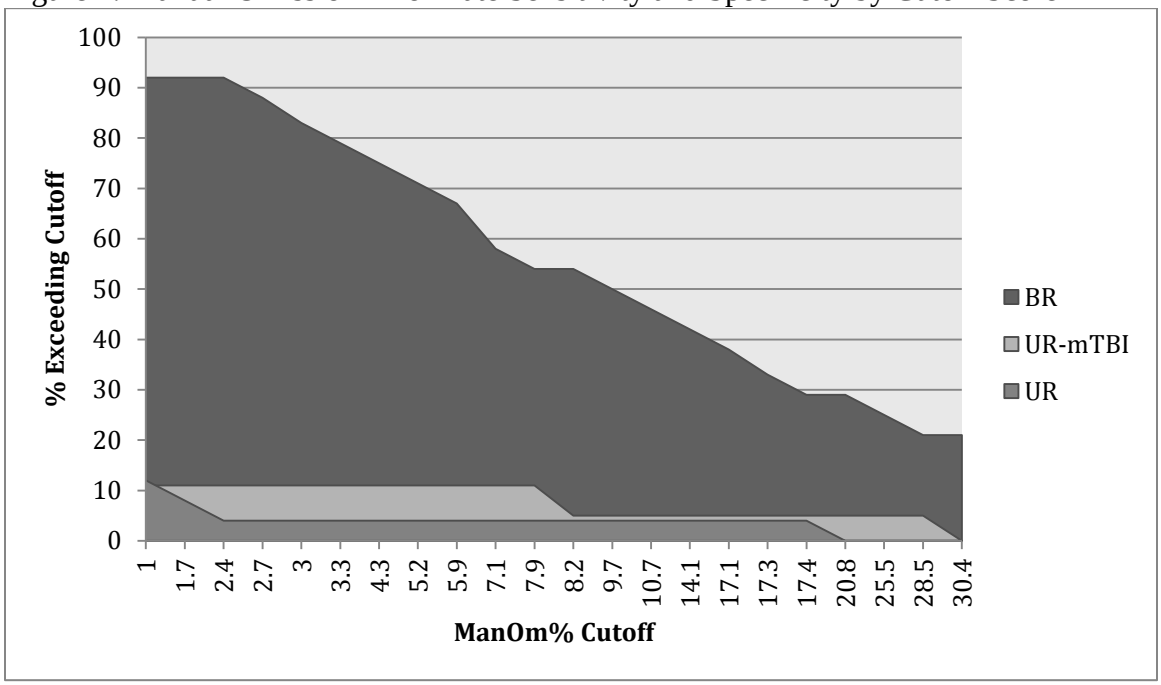
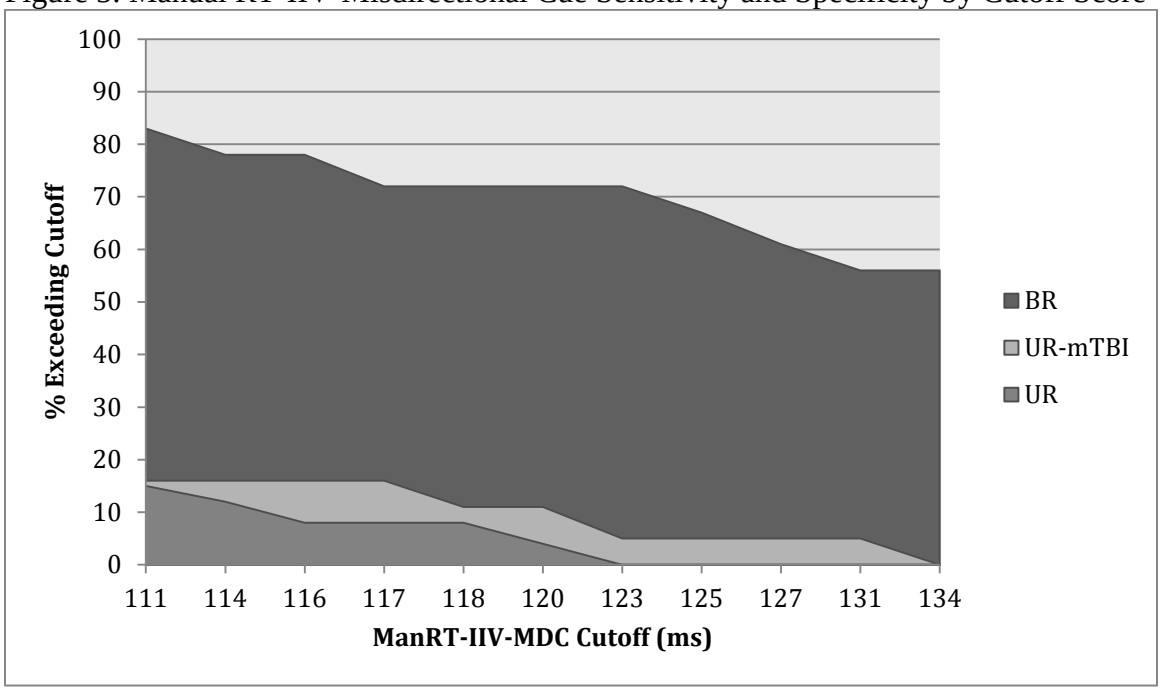


Figure 5: Manual RT-IIV-Misdirectional Cue Sensitivity and Specificity by Cutoff Score

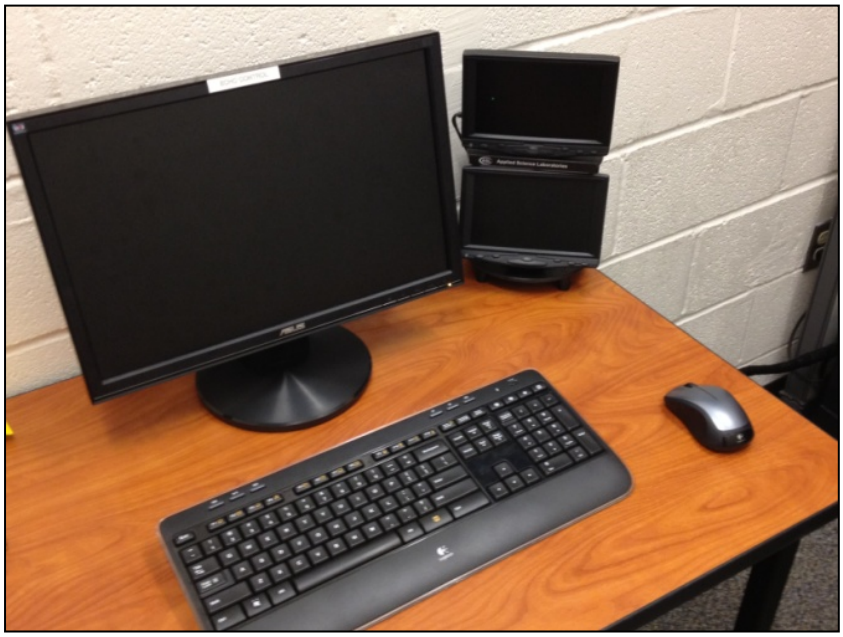


APPENDIX C: SUPPLEMENTAL MATERIALS

Picture 1: Computer Monitor, ASL D6 Eye Tracker, and Cedrus Response Pad

Picture 2: ASL EYE-TRAC 6 Control Unit and Computers

Picture 3: Examiner Station with ASL LCD Monitors



Script 1: Simulator Study Group Assignment Scripts

The following group assignment scripts were adapted from several dissimulation studies (71; 251; 257; 279):

Unbiased Responding (UR) Group Script:

In this study, you will be asked to complete several tests of memory and attention that are often used to evaluate people who sustain head injuries. Your part in this project is to imagine that, six months ago, you were involved in a motor vehicle accident in which you hit your head, but you were not injured at all. Today, you feel normal and unharmed. You have been asked to undergo a routine neuropsychological evaluation to help confirm that there are no problems. Your goal is to prove that your cognitive abilities are normal by *performing as well as you possibly can*.

Please follow instructions carefully and complete all tasks to the best of your ability. Try to do the best you can. Do you understand what you are to do? Do you have any questions?

Biased Responding (BR) Group Script:

In this study, you will be asked to complete several tests of memory and attention that are often used to evaluate people who sustain head injuries. Your part in this project is to take the tests while playing the role of a person who is exaggerating their problems associated with concussions or head injuries. Some individuals who sustain traumatic brain injuries feel normal, or unharmed, following their injury, but they may simulate

injury or exaggerate symptoms to obtain financial rewards. We want to know what this faked performance looks like.

Imagine that, six months ago, you were involved in a motor vehicle accident in which you hit your head. Even though you feel normal today, you know that the amount of money you will receive from your insurance company depends on how badly you were injured. You will try to get extra money by exaggerating your problems on the tests you are about to take. In other words, you are to alter your performance to suggest that your cognitive functioning has been impaired from the head injury you sustained in the accident.

Your goal is to produce the most severely impaired performance you can WITHOUT the examiner knowing that you are faking or pretending. Imagine that if you pretend well enough, you will receive a large sum of money, but if you are caught, you will get nothing. Keep in mind that the deficits you portray must be believable. Major exaggerations, such as remembering absolutely nothing, are easy to detect. The tests you are about to take have ranges of scores associated with brain damage, but also ranges of scores associated with faking bad. Therefore, if you magnify your symptoms too much and they are too obvious, the tests will identify you as someone trying to fake bad rather than someone who is head injured.

Remember, you have to be convincing in your performance. This is going to take some skill on your part. You will have to remind yourself throughout the testing what you are trying to do.

Do you understand what you are to do? Do you have any questions?

Script 2: Simulator Study Debrief Scripts

The following group assignment scripts were adapted from several dissimulation studies (71; 251; 257; 279):

Unbiased Responding (UR) Debrief Script:

Thank you for participating in this study. Your research results will be combined with others who also gave their best effort, and then compared to a group of participants who were asked to perform as if they had sustained a traumatic brain injury. Your participation today will help us identify what elements of our eye tracking measure are affected by test-taking effort, so that we can use this tool most effectively to assess brain functioning after injury. Do you have any questions or concerns?

Biased Responding (BR) Debrief Script:

Thank you for participating in this study. Your research results will be combined with others who also simulated the effects of brain injury, and then compared to a group of participants who were asked to give their best effort. Your participation today will help us identify what elements of our eye tracking measure are affected by test-taking effort, so that we can use this tool most effectively to assess brain functioning after injury. Do you have any questions or concerns?

Measure 1: Simulator Study Feedback Interview

Phase 4 Feedback Interview
(Administered by lab member OTHER THAN the examiner)

Participant ID: _____ Date: _____

1) Participant's group assignment: (circle one) Best effort Simulated TBI

2) "What was your strategy for taking the…

A) Eye tracking measure?"

B) CPT?"

C) VSVT?"

D) Trail A?"

E) Trail B?"

F) Digit Span?"

G) MSVT?"

H) King-Devick Test?"

1

For Simulated TBI group only:

3) "On a scale from 1 to 5, where 5 is the most confident, how confident are you that you performed within the range of a head-injured person on the following tests?"

Eye tracking	CPT	VSVT	Trail A	Trail B	Digit Span	MSVT	K-D Test

4) "Do you have any questions or concerns about the study?"

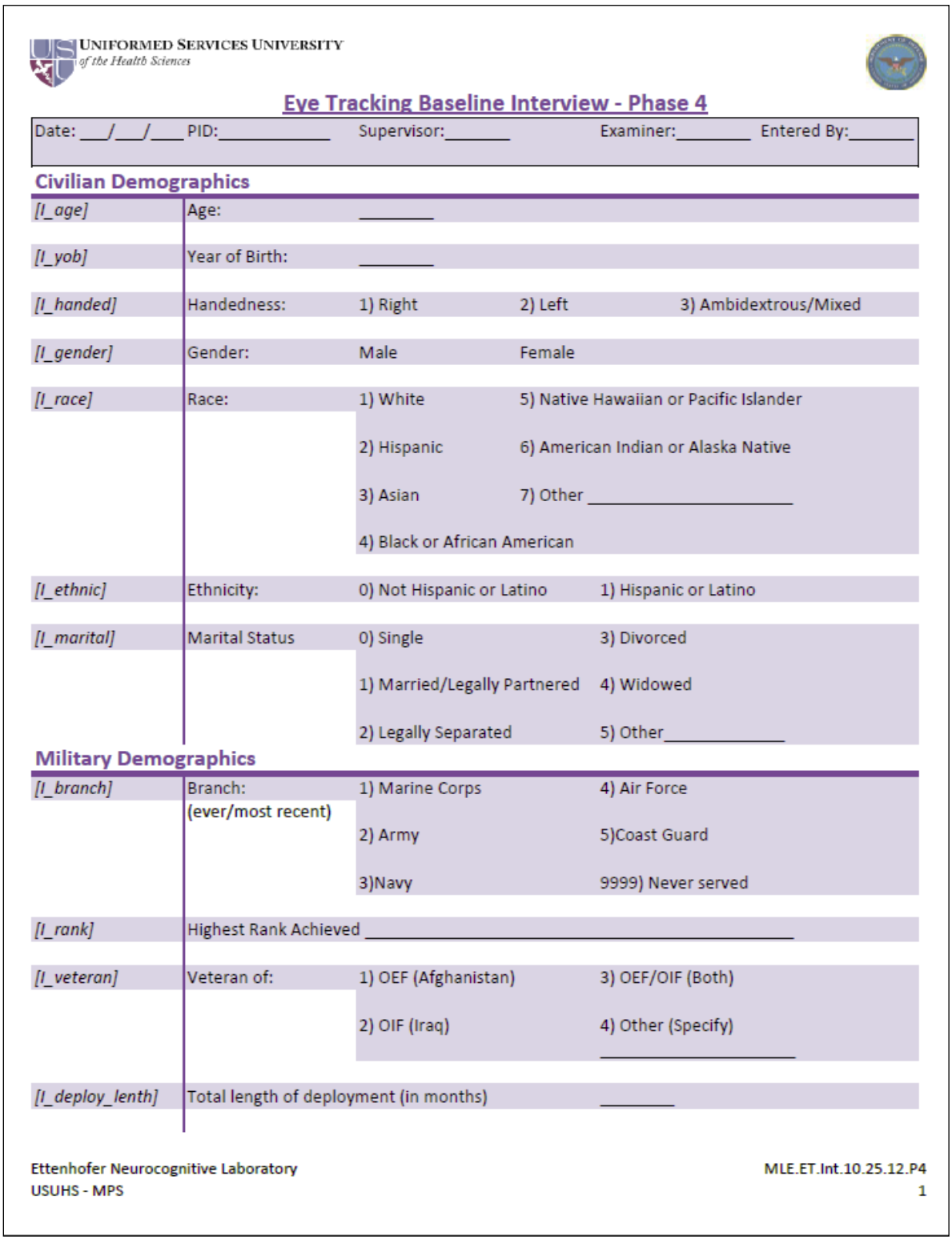
The following questions are meant for the examiner:
(DO NOT tell the participant's group before or after these questions are asked)

1) "What is your best guess of the participant's group assignment?"

(circle one) Best effort Simulated TBI

2) "On a scale from 1 to 5, where 5 is the most confident, how confident are you about your group assignment guess?"

Measure 2: Simulator Study Baseline Interview



UNIFORMED SERVICES UNIVERSITY
of the Health Sciences

Eye Tracking Baseline Interview - Phase 4

Date: ___/___/___ PID:_____ Supervisor:_____ Examiner:_____ Entered By:_____

Civilian Demographics

[I_age]	Age:	_____		
[I_yob]	Year of Birth:	_____		
[I_handed]	Handedness:	1) Right	2) Left	3) Ambidextrous/Mixed
[I_gender]	Gender:	Male	Female	
[I_race]	Race:	1) White	5) Native Hawaiian or Pacific Islander	
		2) Hispanic	6) American Indian or Alaska Native	
		3) Asian	7) Other _____	
		4) Black or African American		
[I_ethnic]	Ethnicity:	0) Not Hispanic or Latino	1) Hispanic or Latino	
[I_marital]	Marital Status	0) Single	3) Divorced	
		1) Married/Legally Partnered	4) Widowed	
		2) Legally Separated	5) Other_____	

Military Demographics

[I_branch]	Branch: (ever/most recent)	1) Marine Corps	4) Air Force
		2) Army	5)Coast Guard
		3)Navy	9999) Never served
[I_rank]	Highest Rank Achieved _____		
[I_veteran]	Veteran of:	1) OEF (Afghanistan)	3) OEF/OIF (Both)
		2) OIF (Iraq)	4) Other (Specify) _____
[I_deploy_lenth]	Total length of deployment (in months)		_____

Ettenhofer Neurocognitive Laboratory
USUHS - MPS

MLE.ET.Int.10.25.12.P4
1

[I_duty]	Duty Status:	0)N/A	1)Reserve	2)Guard	3)Active	4)Separated
[I_duty_date]	Date of separation from duty:			_____		

Educational Background

[I_edu]	Number of years of education: _____ (or number of years before GED, if applicable)		
	GED?	0) No	1) Yes
[I_us_edyrs]	Number of years educated in the US:	_____	
[I_ld]	History of diagnosed learning disability?	0) No	1) Yes
		Specify_____	
[I_adhd]	History of ADHD diagnosis before 18?	0) No	1) Yes

Language Spoken

[I_langprim]	Primary Language Spoken:		
	1) English	2) English as a Second Language	3) Learned both same time
[I_lang1st]	What was your first language (if other than english)?		
	1) Spanish	2) Other (specify) _____	9999) N/A
[I_age_eng]	Age (in years) when first learned English: _____	(zero if from birth)	

Functional Information

[I_employ]	Are you currently employed? (Not including inactive Guard/Reserve)		
	0) No	2) Yes - Part-time (non-military)	
	1) Yes - U.S. Military	3) Yes - Full-time (non-military)	
[I_curred]	Are you currently enrolled in any educational programs? (non-related to any rehab programs)		
	0) No	1) Yes - Part-time	2) Yes - Full-time
[I_disab]	Do you currently receive disability compensation of any kind?		
	0) No	1) Yes - Specify: _____	

UNIFORMED SERVICES UNIVERSITY of the Health Sciences

Participant ID:

Other Medical History

[l_birthcom]	Were there any medical complications at the time of your birth? 0) No 1) Yes - If so, did these affect your health afterward? How?				
[l_birthcom_text]	Specify: _____				
	Do you have a history of any of the following conditions?				
		No	Yes		
[l_tumor]		0	1	Tumor	
[l_thyroid]		0	1	Thyroid Disorder	
[l_cereb_infect]		0	1	Brain Infection	
[l_headache]		0	1	Headache (pre-TBI/no TBI)	
[l_headacheTBI]	N/A	0	1	Headache (post-TBI or circle N/A if no TBI reported)	
[l_vision_prob]		0	1	Vision Problems	
[l_vision_text1]				Specify:_____	
[l_vision_correct]		0	1	Vision Corrected	
[l_vision_text2]				Specify:_____	
[l_seizure]		0	1	Seizures (pre-TBI/no TBI)	
[l_seizureTBI]	N/A	0	1	Seizures (post-TBI or circle N/A if no TBI reported)	
[l_stroke]		0	1	Stroke	
[l_toxic_exp]		0	1	Toxic Exposure	
[l_cvd/_date]		0	1	Cardiovascular Disease	Dx Date: _____
[l_hypertens/_date]		0	1	Diagnosis of Hypertension	Dx Date: _____
[l_heartattack]		0	1	Heart Attack	
[l_cad/_date]		0	1	Coronary Artery Disease	Dx Date: _____
[l_diabetes]		0	1	Diabetes	
[l_smoking/_daily]		0	1	Smoking	Cigs/day: _____
[l_medother_text]		0	1	Other Medical history?	_____

UNIFORMED SERVICES UNIVERSITY
of the Health Sciences

Participant ID:

	Do you take any medications?			
	Medication Name/Purpose	Dosage	Frequency	Hours Since Last Dose
[med1]	1)			
[med2]	2)			
[med3]	3)			
[med4]	4)			
[med5]	5)			
[med6]	6)			
[med7]	7)			
[med8]	8)			
[med9]	9)			
[med10]	10)			

[l_med_psych]	Any psychiatric medications as listed?	0) No	1) Yes
[l_med_pain]	Any pain medications as listed?	0) No	1) Yes
[l_med_sleep]	Any sleep medications as listed?	0) No	1) Yes
[l_sleep_lastnight]	How many hours of sleep did you get last night?		____
[l_sleep_typical]	How many hours of sleep per night has been typical for you in the last month?		____
[l_fatigue]	On a scale of one to ten (ten being the most tired), what is your current level of fatigue?		

1 2 3 4 5 6 7 8 9 10

Substances

	Have you had any of these within the last 12 hours? No	Yes					
[I_alcohol]	0	1	Alcohol	# Drinks _____	Hours since use _____		
[I_nicotine]	0	1	Nicotine	# Cigs _____	Hours since use _____		
[I_caffeine]	0	1	Caffeine	# Drinks _____	Hours since use _____		

Individuals who have had more than 3 drinks within the last 12 hours, or any drinks within the last 4 hours should NOT be tested (reschedule testing)

Notes:_____

_____

_____

_____

Measure 3: Parent Study TBI Cohort Baseline Interview

UNIFORMED SERVICES UNIVERSITY
of the Health Sciences

Eye Tracking Baseline Interview - Phase 3

Date: ___/___/___ PID:_____ Examiner:_____ Entered By:_____

General Demographics

[l_age]	Age:	_____		
[l_yob]	Year of Birth:	_____		
[l_handed]	Handedness:	1) Right	2) Left	3) Ambidextrous/Mixed
[l_gender]	Gender:	1) Male	2) Female	
[l_race]	Race:	1) White	5) Native Hawaiian or Pacific Islander	
		2) Hispanic	6) American Indian or Alaska Native	
[l_race_text]		3) Asian	7) Other _____	
		4) Black or African American		
[l_ethnic]	Ethnicity:	0) Not Hispanic or Latino	1) Hispanic or Latino	
[l_marital]	Marital Status:	0) Single	3) Divorced	
		1) Married/Legally Partnered	4) Widowed	
[l_marital_text]		2) Legally Separated	5) Other_____	

Military Demographics

[l_branch]	Branch: (ever/most recent)	1) Marine Corps	4) Air Force	0) Never served
		2) Army	5)Coast Guard	
[l_branch_txt]		3)Navy	6) Other _____	
[l_rank]	Highest Rank Achieved: _____			
[l_veteran]	Veteran of:	1) OEF (Afghanistan)	3) OEF/OIF (Both)	0) None
		2) OIF (Iraq)	4) Other (Specify) _____	

Ettenhofer Neurocognitive Laboratory
USUHS - MPS

MLE.ET.Int.06.15.12.P3
1

[l_deploy_length]	Total length of all deployments (in months): _____	0) None
[l_combat_length]	Total length of deployment in a combat zone (in months): _____	0) None
[l_duty]	Duty Status: 9999)N/A 1)Reserve 2)Guard 3)Active 4)Separated	
[l_duty_date]	Date of separation from duty: ____/___/____	
[l_sep_text]	Conditions of separation: _____	

Educational Background

[l_edu]	Highest level of education: _____ (or number of years before GED, if applicable)	
[l_ged]	GED (versus high school diploma): 0) No 1) Yes	
[l_us_edyrs]	Number of years educated in the US: _____	
[l_ld]	History of diagnosed learning disability: 0) No 1) Yes	
[l_ldt_text]	Specify_____	
[l_adhd]	History of ADHD diagnosis before 18: 0) No 1) Yes	

Language Spoken

[l_langprim]	Was English your second language?	
	1) No 2) Yes 3) Learned both languages at the same time	
[l_1stlang]	What was your first language (if other than English)?	
[l_1stlang_text]	1) Spanish 2) Other (specify) _____ 9999) N/A	
[l_age_eng]	Age (in years) when first learned English: _____ (zero if from birth)	

Functional Information

[l_employ]	Are you currently employed? (Not including inactive Guard/Reserve)	
	0) No 2) Yes - Part-time (non-military)	
	1) Yes - U.S. Military 3) Yes - Full-time (non-military)	
[l_curred]	Are you currently enrolled in any educational programs? (non-related to any rehab programs)	
	0) No 1) Yes - Part-time 2) Yes - Full-time	

[l_disab] Do you currently receive disability compensation of any kind?

[l_disab_text] 0) No 1) Yes - Specify: _____

Do you currently receive help (from family, paid helpers, etc) or share responsibilities with somoeone else for any of the following activities?

Self-care	Financial	Housework	Shopping
(washing, dressing) *[l_currselfcare]*	*[l_currfinancial]*	(laundry, cleaning) *[l_currhousework]*	*[l_currshopping]*
0) No help	0) No help	0) No help	0) No help
1) Some	1) Some	1) Some	1) Some
2) Majority/Total	2) Majority/Total	2) Majority/Total	2) Majority/Total
Highest Previous	Highest Previous	Highest Previous	Highest Previous
[l_highselfcare]	*[l_highfinancial]*	*[l_highhousework]*	*[l_highshopping]*
0) No help	0) No help	0) No help	0) No help
1) Some	1) Some	1) Some	1) Some
2) Majority/Total	2) Majority/Total	2) Majority/Total	2) Majority/Total

[l_hobbies] On a scale of 1-10 (with 10 being the highest), how satisfied are you with your current involvement in personal hobbies and interests?

1 2 3 4 5 6 7 8 9 10

[l_social] On a scale of 1-10 (with 10 being the highest), how satisfied are you with your current involvement in social activities?

1 2 3 4 5 6 7 8 9 10

Most Recent Traumatic Brain Injury History

[l_num_tbi] Have you ever lost consciousness or forgotten information as a result of an explosion or a blow to the head? If so, **how many times** has this occurred throughout the course of your lifetime? (note: this requires LOC or PTA)

_____

[l_datetbi1] When was the most recent incident?

____/____/____ Estimates are OK N/A= 99/99/9999

UNIFORMED SERVICES UNIVERSITY
of the Health Sciences

Participant ID:

[l_tbi1_text] Brief description of the MOST RECENT incident: _____

_____

_____

[l_tbi1injury] Was any part of your body injured as a result of this incident, other than your head?

9999) N/A 0) No 1) Yes Describe: _____

[l_tbi1injury_text] _____

[l_mil1] Was this incident during a military deployment?

1) Before any deployment
2) During deployment
3) After deployment
4) Between deployments
9999) N/A

[l_activity1] What was the cause of the incident? (choose only one closest answer, or "other")

1) Motor vehicle crash
2) Sports/Recreation
3) Assault (person on person)
4) Hit by (moving) projectile
5) In vehicle during blast
6) Accidental fall
7) On foot during blast
8) Other (specify)

[l_activity1_text] _____

9999) N/A (no TBI reported)

How was this head injury acquired?

Obj hit head	Head hit object	Blast injury	Projectile	Other
[l_force1ohh]	*[l_force1hho]*	*[l_force1blast]*	*[l_force1proj]*	*[l_force1other]*
0) No	0) No	0) No	0) No	0) No
1) Yes	1) Yes	1) Yes	1) Yes	1) Yes
9999) N/A	9999) N/A	9999) N/A	9999) N/A	9999) N/A

[l_force1_text] If other, then specify: _____

[l_helmet_tbi1]	Were you wearing a helmet at the time of the incident? 0) No 1) Yes 9999) N/A (No TBI reported)
[l_loctime1] **BE AS ACCURATE AS POSSIBLE**	How long were you unconscious? (according to incident report, if available) **DO THE MATH! CONVERT TO MINS!** _____ **Minutes** _____ (Hours) _____ (Days) 7777) Unknown (due to PTA) 8888) Yes LOC, but insufficient info 9999) N/A (No TBI)
[l_locmin1]	LOC Minimum - troubleshoot as needed *(e.g. if injured in Iraq and woke up in Germany, implies LOC>6 hours)* **DO THE MATH! CONVERT TO MINS!** _____ **Minutes** _____ (Hours) _____ (Days) 9999) N/A (No TBI)
[l_locmax1]	LOC Maximum - troubleshoot as needed **DO THE MATH! CONVERT TO MINS!** _____ **Minutes** _____ (Hours) _____ (Days) 9999) N/A (No TBI)
[l_ptatime1] **BE AS ACCURATE AS POSSIBLE**	Is any time "missing" from your memory after the injury? If so, how much? (use PTA according to incident report, if available) **DO THE MATH! CONVERT TO MINS!** _____ **Minutes** _____ (Hours) _____ (Days) 8888) Yes PTA, but insufficient info 9999) N/A (No TBI)
[l_ptamin1]	PTA Minimum - troubleshoot as needed **DO THE MATH! CONVERT TO MINS!** _____ **Minutes** _____ (Hours) _____ (Days) 9999) N/A (No TBI)
[l_ptamax1]	PTA Maximum - troubleshoot as needed **DO THE MATH! CONVERT TO MINS!** _____ **Minutes** _____ (Hours) _____ (Days) 9999) N/A (No TBI)

[l_punct1]	Was your skull punctured?	0) No	1) Yes	9999) N/A (no TBI reported)
[l_punct1_text]	If so, specify: _____			
[l_fract1]	Was your skull fractured?	0) No	1) Yes	9999) N/A (no TBI reported)
[l_fract1_text]	If so, specify: _____			
[l_tx1]	Did you receive treatment or evaluation shortly after the time of the injury?			
	0) No, none	1) Yes		9999) N/A (no TBI reported)
[l_txh1]	Were you hospitalized for this incident? If yes, how long was the hospitalization? Note: hospitalization means inpatient admit for >24 hours			
	_____ Days	0) No Hospitalization		
	8888) Yes hospitalization, but duration unknown			9999) N/A (no TBI reported)
[l_lgl1]	Have you ever been involved in litigation related to this injury			
	0) No	1) Yes		9999) N/A (no TBI reported)

Most Severe Traumatic Brain Injury History

[l_date_tbi2]	When was your **most severe** head injury? DO NOT CODE AGAIN FOR SAME INJURY AS ABOVE Note: most severe according to PTA/LOC (etc.; follow up as needed)
	____/____/____ 99/99/9999) N/A
	Brief description of most severe head injury incident:
[l_tbi2_text]	_____
	_____
	_____
[l_tbi2injury]	Was any part of your body injured as a result of this incident, other than your head?
	9999) N/A 0) No 1) Yes Describe: _____
[l_tbi2injury_text]	_____

UNIFORMED SERVICES UNIVERSITY of the Health Sciences

Participant ID:

[l_mil2]

Was this incident during a military deployment?

1) Before any deployment
2) During deployment
3) After deployment
4) Between deployments
9999) N/A

[l_activity2]

[l_activity2_text]

What was the cause of the incident? (choose only one closest answer, or "other")

1) Motor vehicle crash
2) Sports/Recreation
3) Assault (person on person)
4) Hit by (moving) projectile
5) In vehicle during blast
6) Accidental fall
7) On foot during blast
8) Other - Specify: _____
9999) N/A (no TBI reported)

How was this head injury acquired?

Obj hit head	Head hit object	Blast injury	Penetrated Skull	Other
[l_force2ohh]	*[l_force2hho]*	*[l_force2blast]*	*[l_force2proj]*	*[l_force2other]*
0) No	0) No	0) No	0) No	0) No
1) Yes	1) Yes	1) Yes	1) Yes	1) Yes
9999)	9999)	9999)	9999)	9999) N/A

[l_force2_text]

If other, then specify:

_____

[l_helmet_tbi2]

Were you wearing a helmet at the time of the incident?

0) No 1) Yes 9999) N/A (No TBI reported)

[l_loctime2]

BE AS ACCURATE AS POSSIBLE

How long were you unconscious? (refer to incident report, if available) **DO THE MATH! CONVERT TO MINS!**

_____ **Minutes** _____ (Hours) _____ (Days)

7777) Unknown (due to PTA) 8888) Yes LOC, but insufficient info 9999) N/A (No TBI)

[l_locmin2]	LOC Minimum - troubleshoot as needed *(e.g. if injured in Iraq and woke up in Germany, implies LOC>6 hours)* _____ **Minutes** _____(Hours) _____ (Days) 9999) N/A (No TBI)	**DO THE MATH! CONVERT TO MINS!**
[l_locmax2]	LOC Maximum - troubleshoot as needed _____ **Minutes** _____(Hours) _____ (Days) 9999) N/A (No TBI)	**DO THE MATH! CONVERT TO MINS!**
[l_ptatime2] **BE AS ACCURATE AS POSSIBLE**	Is any time "missing" from your memory after the injury? If so, how much? (use PTA according to incident report, if available) _____ **Minutes** _____(Hours) _____ (Days) 8888) Yes LOC, but insufficient info 9999) N/A (No TBI)	**DO THE MATH! CONVERT TO MINS!**
[l_ptamin2]	PTA Minimum - troubleshoot as needed _____ **Minutes** _____(Hours) _____ (Days) 9999) N/A (No TBI)	**DO THE MATH! CONVERT TO MINS!**
[l_ptamax2]	PTA Maximum - troubleshoot as needed _____ **Minutes** _____(Hours) _____ (Days) 9999) N/A (No TBI)	**DO THE MATH! CONVERT TO MINS!**
[l_punct2]	Was your skull punctured? 0) No 1) Yes 9999) N/A (no TBI reported)	
[l_punct2_text]	If so, specify: _____	
[l_fract2]	Was your skull fractured? 0) No 1) Yes 9999) N/A (no TBI reported)	
[l_fract2_text]	If so, specify: _____	

[l_tx2]	Did you receive treatment or evaluation shortly after the time of the injury? 0) No, none 1) Yes 9999) N/A (no TBI reported)
[l_txh2]	Were you hospitalized for this incident? If yes, how long was the hospitalization? Note: hospitalization means inpatient admit for >24 hours _____ Days 0) No Hospitalization 8888) Yes hospitalization, but duration unknown 9999) N/A (no TBI reported)
[l_lgl2]	Have you ever been involved in litigation related to this injury 0) No 1) Yes 9999) N/A (no TBI reported)

Additional TBI Incidents

[l_blast_noTBI] **ESTIMATES OK**	How many times were you exposed to a blast where you felt the pressure wave on your body that DID NOT result in you losing consciousness, having any memory loss, or feeling confused or disoriented? _____
[l_blastTBI] **ESTIMATES OK**	How many times were you exposed to a blast that DID result in you losing consciousness, feeling confused/disoriented, or having any memory loss? _____
[l_num_aoc] **ESTIMATES OK**	Have you ever been confused or disoriented after an explosion or a blow to the head, but did NOT lose consciousness or experience memory loss? If so, **how many times** has this occurred throughout the course of your life? (Note: this is the number of times experiencing AOC without LOC or PTA)
[l_numaoc_text]	_____ If yes, specify:_____

Other Medical History

[l_birthcom]	Were there any medical complications at the time of your birth? 0) No 1) Yes - If so, did these affect your health afterward? How?
[l_birthcom_text]	Specify: _____

UNIFORMED SERVICES UNIVERSITY of the Health Sciences

Participant ID:

	Do you have a history of any of the following conditions?	No	Yes	
[l_tumor]		0	1	Tumor
[l_thyroid]		0	1	Thyroid Disorder
[l_cereb_infect]		0	1	Brain Infection
[l_headache]		0	1	Headache (pre-TBI/no TBI)
[l_headacheTBI]	N/A	0	1	Headache (post-TBI or circle N/A if no TBI reported)
[l_vision_prob]		0	1	Vision Problems
[l_vision_text1]				Specify:_____
[l_vision_correct]		0	1	Vision Corrected
[l_vision_text2]				Specify:_____
[l_seizure]		0	1	Seizures (pre-TBI/no TBI)
[l_seizureTBI]	N/A	0	1	Seizures (post-TBI or circle N/A if no TBI reported)
[l_stroke]		0	1	Stroke
[l_toxic_exp]		0	1	Toxic Exposure
[l_cvd/_date]		0	1	Cardiovascular Disease Dx Date: _____
[l_hypertens/_date]		0	1	Diagnosis of Hypertension Dx Date: _____
[l_heartattack]		0	1	Heart Attack
[l_cad/_date]		0	1	Coronary Artery Disease Dx Date: _____
[l_diabetes]		0	1	Diabetes
[l_smoking/_daily]		0	1	Smoking Cigs/day: _____
[l_medother_text]		0	1	Other Medical history? _____

UNIFORMED SERVICES UNIVERSITY of the Health Sciences

Participant ID:

	Have you ever been **diagnosed** with:	No	Yes	
[l_depression]		0	1	Clinical Depression
[l_ptsd]		0	1	PTSD
[l_anxiety]		0	1	Anxiety (not related to PTSD)
[l_anxiety_text]				Specify:_____
[l_psychotic]		0	1	Psychotic Disorders (e.g. Schizophrenia)
[l_psych_other]		0	1	Other - Specify:_____
[l_psychother_text]				_____

[l_psychtx]	Are you currently involved in a behavioral health program?
[l_psychtx_text]	0) No 1) Yes Specify:_____
[l_cogtx]	Are you currently in a cognitive rehabilitation program? (follow up if they aren't sure)
[l_cogtx_text]	0) No 1) Yes Specify:_____

Do you take any medications?

	Medication Name and Purpose	Dosage	Frequency	Hours Since Last Dose
[med1]	1)			
[med2]	2)			
[med3]	3)			
[med4]	4)			
[med5]	5)			
[med6]	6)			
[med7]	7)			
[med8]	8)			
[med9]	9)			
[med10]	10)			

[l_med_psych]	Any psychiatric medications as listed?	0) No	1) Yes
[l_med_pain]	Any pain medications as listed?	0) No	1) Yes
[l_med_sleep]	Any sleep medications as listed?	0) No	1) Yes
[l_sleep_lastnight]	How many hours of sleep did you get last night?		____
[l_sleep_typical]	How many hours of sleep per night has been typical for you in the last month?		____
[l_fatigue]	On a scale of one to ten (ten being the most tired), what is your current level of fatigue?		

1 2 3 4 5 6 7 8 9 10

Other Substances

	Have you had any of these within the last 12 hours? No	Yes				
[l_alcohol]	0	1	Alcohol # Drinks	_____	Hours since use	_____
[l_nicotine]	0	1	Nicotine # Cigs	_____	Hours since use	_____
[l_caffeine]	0	1	Caffeine # Drinks	_____	Hours since use	_____

Individuals who have had more than 3 alcoholic drinks within the last 12 hours, or any drinks within the last 4 hours should NOT be tested (reschedule testing)

GENERAL NOTES: _____

_____

_____

_____

_____

Measure 4: Head Injury Knowledge Scale Version B (HIKS B)

Experiences Following Brain Injury: Version B

Having a brain injury can lead to a range of changes in a person's everyday abilities. We are interested in what changes you believe are likely to occur after someone has had a brain injury.

For each item, please circle TRUE if you think this would occur often or most of the time or FALSE if you think this would rarely or never occur following a brain injury.

	Often or most of the time	Rarely or never
Have difficulty recognising faces of family members	True	False
Have trouble concentrating on more than one task at a time	True	False
Less movement or coordination down one side of the body	True	False
Have trouble remembering major events from childhood	True	False
Think and act as if you were a different person altogether	True	False
See things or images that are not really there	True	False
Have trouble finding the right words in conversation	True	False
Experience a reduced or loss of sense of smell	True	False
Have trouble saying the letters of the alphabet	True	False
Become frustrated more easily	True	False
Forget your own first name or surname	True	False
Feel tired more easily	True	False
Have trouble remembering how to get dressed	True	False
Say things without thinking them through first	True	False
Feel pins and needles in the face and both arms	True	False
Plan to do things in the future but do not follow them through	True	False
Become upset and yell for no reason	True	False
Have trouble remembering details of recent conversations	True	False

Administrative Document 1: Simulator Study Informed Consent Form

UNIFORMED SERVICES UNIVERSITY
OF THE HEALTH SCIENCES

INFORMED CONSENT FORM
RESEARCH STUDY

Eye Tracking Indicators of Neurocognitive Status after Traumatic Brain Injury – Phase 4

INTRODUCTION

You are being asked to take part in a research study. Before you decide if you want to be in the study, you need to understand its risks and benefits so that you can make an informed decision. This is known as informed consent.

This consent form provides information about the research study which has been explained to you. Once you understand what it involves, you will be asked to tell the researcher if you want to take part in it. Your decision to take part in the study is entirely voluntary. This means that you are free to choose whether or not you want to be a research subject.

DESCRIPTION OF THE RESEARCH AND ITS PURPOSE

The purpose of this study is to develop and test an eye-tracking tool to accurately diagnose traumatic brain injury. This tool works by watching your eyes as you complete tasks on a computer.

In this phase of the experiment, you will be randomly assigned to one of two groups, each with a different set of instructions for how to complete a series of tasks. You will be asked to follow the instructions to the best of your ability. All participants will complete computer tasks while your eye movements are tracked. You will also complete a number of other thinking tasks.

This study is being conducted using funds from the Uniformed Services University of the Health Sciences (USUHS).

The principal investigator for this study is: Mark L. Ettenhofer, Ph.D.
Department of Medical and Clinical Psychology
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814-4712
301-295-3279

Eligibility to Participate:

You are being asked to be in this study because you were previously determined to be eligible during the pre-screening procedure. During the pre-screening procedure, it was determined that you are over the age of 18, and you do not have a medical condition that would be expected to affect your eye or brain functioning or your use of your hands.

If you are Active Duty Military or a civilian federal employee, it is also required that you provide a signed Statement of Approval for Participation in Research. Active Duty Military personnel must have this approval signed by their supervisor and the Brigade Commander; Federal Civilians must have this signed by their Supervisor before any research participation.

STUDY PROCEDURE:

Your participation in this study will require **1** study visit that will last about **1.5 hours.**

If you agree to participate, you will sign this consent form after it has been explained to you and before any study-related procedures take place.

We will collect personal information about you (your name, address, phone number, and email address) so that we may compensate you for participation in this study. You will be randomly assigned to one of two groups, each with a different set of instructions for how to complete a series of tasks. You will be asked to follow the instructions to the best of your ability. You will complete tasks, procedures, and questionnaires during this time to measure your thinking (e.g., attention, reaction time, memory). You may refuse to answer any questions that make you feel uncomfortable, or withdraw from the study at any time. You will also complete a series of computer tasks, about 45 minutes in duration, during which your eyes will be tracked by a camera. The computer will record your eye movements while you complete the tasks.

POSSIBLE BENEFITS

The information researchers get from this study may help others in the future. You might not personally benefit from being in this study.

COMPENSATION

If you are active duty military or a federal employee, you are not eligible for compensation. Otherwise, you will be compensated $30 for your participation, paid by check after your participation visit is complete. The check will be mailed if requested.

POSSIBLE RISKS

There are no known or expected risks for participating in this study, but you could have side effects that we do not expect or know to watch for now. Call the principal investigator if you have any symptoms or problems.

There is a risk that one or more of these questions or tasks might make you upset or uncomfortable. If this happens, remember that you will not need to respond to any questions, complete any tasks, or follow any instructions that make you feel upset or uncomfortable. You may also discontinue participation at any time without penalty.

Referrals

If we feel it is needed or you request it, we will provide you with referrals to a mental health care provider for evaluation or treatment at your option and your expense. These referrals may be provided up to one week from your visit if the principal investigator judges that you may benefit from these services based upon evidence of mental health difficulties. However, this study is not intended to diagnose or treat any conditions. Non-referral does not imply the absence of a mental health condition.

RIGHT TO WITHDRAW FROM THE STUDY

You may decide to stop taking part in this study at any time. This will not affect your relationship with USUHS in any way. You can agree to be in the study now and change your mind later. If you decide to withdraw, you may do so in person or over the phone by calling Dr. Ettenhofer at 301-295-3279. However, phone calls and any formal withdrawals must be accompanied by a written, signed request including your full printed name, and sent to Dr. Ettenhofer at the address listed above. Your participation may also be discontinued by study personnel for reasons including, but not limited to, your potential difficulty following study procedures. If requested, we will also destroy any information we have collected about you.

PRIVACY AND CONFIDENTIALITY

All information you provide as part of this study will be confidential and will be protected to the fullest extent provided by law. Your records related to this study will be accessible to those persons directly involved in conducting this study and members of the Uniformed Services University of the Health Sciences Institutional Review Board (IRB), which provides oversight for protection of human research volunteers. In addition, the Institutional Review Board at USUHS and other federal agencies who help protect people who are involved in research studies, may need to see the information you give us. Other than those groups, records from this study will be kept private to the fullest extent of the law. Scientific reports that come out of this study may use the information you have provided, but these reports will not use your name or identify you in any way.

Records of your participation in this study may only be disclosed in accordance with federal law, including the Federal Privacy Act, 5 U.S.C.552a, and its implementing regulations. Confidentiality of your records will be protected to the extent possible under existing regulations and laws but cannot be guaranteed. Complete confidentiality cannot be promised, particularly for military personnel, because information bearing on your health may be required to be reported to appropriate medical or command authorities.

Personal contact information may be retained for the purposes of completing this study and to notify you of future studies and assess your interest in participation. You will only be contacted regarding your current participation and future studies. Optionally, you may choose to not be contacted for future studies by notifying study personnel of your decision.

RECOURSE IN THE EVENT OF INJURY

This study should not entail any physical or mental risk beyond those described above. We do not expect complications to occur, but if, for any reason, you feel that continuing this study would constitute a hardship for you, we will end your participation in the study.

In the event of a medical emergency while participating in this study or medical treatment required as a result of your participation in this study, you may receive emergency treatment in the facility you are in or a nearby Department of Defense (military) medical facility (hospital or clinic). Treatment/care will be provided even if you are not eligible to receive such care. Care will be continued until the medical doctor treating you decides that you are out of immediate danger. If you are not entitled to care in a military facility, you may be transferred to a private civilian hospital. The attending doctor or member of the hospital staff will go over the transfer decision with you before it happens. The military will bill your health insurance for health care you receive which is not part of the study. You will not be personally billed and you WILL NOT be expected to pay for medical care at our hospitals. If you are required to pay a deductible you may make a claim for reimbursement through the Uniformed Services University Office of General Counsel. In case you need additional care following discharge from the military hospital or clinic, a military health care professional will decide whether your need for care is directly related to being in the study. If your need for care is related to the study, the military may offer you limited health care at its medical facilities. This additional care is not automatic.

If at any time you believe you have suffered an injury or illness as a result of participating in this research project, you should contact the Office of Research at the Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799 at (301) 295-3303. This office can review the matter with you, can provide information about your rights as a subject, and may be able to identify resources available to you. If you believe the government or one of the government's employees (such as a military doctor) has injured you, a claim for damages (money) against the federal government (including the military) may be filed under the Federal Torts Claims Act. Information about judicial avenues of compensation is available from the University's General Counsel at (301) 295-3028.

IF YOU HAVE QUESTIONS OR CONCERNS

If you have questions about this research, you should contact Dr. Mark Ettenhofer, the person in charge of the study. His phone number at USUHS is 301-295-3279. Even in the evening or on weekends, you can leave a message at that number. If you have questions about your rights as a research subject, you should call the Director, Human Research Protections Program at USUHS at (301) 295-9534. He is your representative and has no connection to the researcher conducting this study.

By signing this form you are agreeing that this study has been explained to you, that you understood that explanation, and that you want to take part in this research.

_____ _____

Subject Date of signature

_____ _____

Witness Date of signature

I certify that the research study has been explained to the above individuals, by me or my research staff, and that the individual understands the nature and purpose, the possible risks and benefits associated with taking part in this research study. Any questions that have been raised have been answered.

_____ _____

Investigator Date of signature

Administrative Document 2: Parent Study TBI Cohort Informed Consent Form

UNIFORMED SERVICES UNIVERSITY
OF THE HEALTH SCIENCES

INFORMED CONSENT FORM
RESEARCH STUDY

Eye Tracking Indicators of Neurocognitive Status after Traumatic Brain Injury – Phase 3

INTRODUCTION

You are being asked to take part in a research study. Before you decide if you want to be in the study, you need to understand its risks and benefits so that you can make an informed decision. This is known as informed consent.

This consent form provides information about the research study which has been explained to you. Once you understand what it involves, you will be asked to tell the researcher if you want to take part in it. Your decision to take part in the study is entirely voluntary. This means that you are free to choose whether or not you want to be a research subject.

DESCRIPTION OF THE RESEARCH AND ITS PURPOSE

The purpose of this study is to develop and test an eye-tracking tool to accurately diagnose traumatic brain injury. This tool works by watching your eyes as you complete tasks on a computer.

In this phase of the experiment, you will complete computer tasks while your eye movements are tracked. You will also complete a number of other thinking tasks and provide information about your psychological functioning. We will compare your results to the results of other people without concussions or brain injuries in order to determine which tests tend to show different results between groups.

This study is being conducted using funds from the Uniformed Services University of the Health Sciences (USUHS).

The principal investigator for this study is:

Mark L. Ettenhofer, Ph.D.
Department of Medical and Clinical Psychology
Uniformed Services University of the Health Sciences
4301 Jones Bridge Road
Bethesda, MD 20814-4712
301-295-3279

USUHS IRB APPROVED
11 OCT 12
Expires: 10 OCT 13

Eligibility to Participate:

You are being asked to be in this study because you were previously determined to be eligible during the pre-screening procedure. During the pre-screening procedure, it was determined that you are over the age of 18, you have a history of concussion or brain injury, and you do not have

any other medical condition that would be expected to affect your eye or brain functioning or your use of your hands.

If you are Active Duty Military or a civilian federal employee it is also required that you provide a signed Statement of Approval for Participation in Research. Active Duty Military personnel must have this approval signed by their supervisor and the Brigade Commander; Federal Civilians must have this signed by their Supervisor before any research participation.

STUDY PROCEDURE:

Your participation in this study will require about **2.5 hours** during a single visit. We will also call you 6 and 12 months from now to conduct follow-up interviews of about **10 minutes** each.

Before you decide whether or not you agree to participate, your ability to provide informed consent will be evaluated. If you are determined not to have the ability to provide informed consent, you will not be permitted to participate in the study. If you agree to participate, you will sign this consent form after it has been explained to you and before any study related procedures take place. All Active Duty and Veteran Military participants will also be asked if they are willing to provide written permission for separate release of information (ROI) of prior Armed Services Vocational Aptitude Battery (ASVAB) test results (as available) for inclusion in the study database. This information will allow study personnel to make comparisons between your performance on that previous test and the tests you will complete during this study. Declining to provide permission for release of ASVAB results will not result in exclusion from the study or affect other aspects of study participation.

Visit 1: (2.5 hours)

We will collect personal information about you (your name, address, phone number, and the name and phone number of two people you know) so that we are able to contact you to complete your telephone follow-up. We will review your medical history. You will complete tasks, procedures, and questionnaires during this time to measure your thinking (e.g., attention, reaction time, memory) and your psychological functioning. You may refuse to answer any questions that make you feel uncomfortable. You will also complete a series of computer tasks, about 30 minutes in duration, during which your eyes will be tracked by a camera. The computer will record your eye movements while you complete the tasks.

Telephone Follow Ups: (10 minutes)

This study will also include a follow up interview with you at 6 and 12 months after the initial visit. You will be asked to complete a small survey over the phone that should take approximately 10-15 minutes. We will ask about your health and your general life functioning.

POSSIBLE BENEFITS

The information researchers get from this study may help others in the future. You might not personally benefit from being in this study.

COMPENSATION

If you are active duty military or a federal employee, you are not eligible for compensation. Otherwise, you will be compensated $40 for your participation, paid by check after your first visit (e.g., prior to telephone follow-ups).

POSSIBLE RISKS

There are no known or expected risks for participating in this study, but you could have side effects that we do not expect or know to watch for now. Call the principal investigator if you have any symptoms or problems.

There is a risk that one or more of these questions or tasks might make you upset or uncomfortable. If this happens, remember that you will not need to respond to any questions or complete any tasks that make you feel upset or uncomfortable. You may also discontinue participation at any time without penalty.

Referrals

If we feel it is needed or you request it, we will provide you with referrals to a mental health care provider for evaluation or treatment at your option and your expense. These referrals may be provided up to one week from your visit if the principal investigator judges that you may benefit from these services based upon evidence of mental health difficulties. However, this study is not intended to diagnose or treat any conditions. Non-referral does not imply the absence of a mental health condition.

RIGHT TO WITHDRAW FROM THE STUDY

You may decide to stop taking part in this study at any time. This will not affect your relationship with USUHS in any way. You can agree to be in the study now and change your mind later. If you decide to withdraw, you may do so in person or over the phone by calling Dr. Ettenhofer at 301-295-3279. However, phone calls and any formal withdrawals must be accompanied by a written, signed request including your full printed name, and sent to Dr. Ettenhofer at the address listed above. Your participation may also be discontinued by study personnel for reasons including, but not limited to, your potential difficulty following study procedures. If requested, we will also destroy any information we have collected about you.

PRIVACY AND CONFIDENTIALITY

All information you provide as part of this study will be confidential and will be protected to the fullest extent provided by law. Your records related to this study will be accessible to those persons directly involved in conducting this study and members of the Uniformed Services University of the Health Sciences Institutional Review Board (IRB), which provides oversight for protection of human research volunteers. In addition, the Institutional Review Board at USUHS and other federal agencies who help protect people who are involved in research studies, may need to see the information you give us. Other than those groups, records from this study will be kept private to the fullest extent of the law. Scientific reports that come out of this study

USUHS IRB APPROVED
Expires: 10 OCT 13

may use the information you have provided, but these reports will not use your name or identify you in any way.

Records of your participation in this study may only be disclosed in accordance with federal law, including the Federal Privacy Act, 5 U.S.C.552a, and its implementing regulations. Confidentiality of your records will be protected to the extent possible under existing regulations and laws but cannot be guaranteed. Complete confidentiality cannot be promised, particularly for military personnel, because information bearing on your health may be required to be reported to appropriate medical or command authorities.

Personal contact information may be retained for the purposes of completing this study and to notify you of future studies and assess your interest in participation. You will only be contacted regarding your current participation and future studies. Optionally, you may choose to not be contacted for future studies by notifying study personnel of your decision.

RECOURSE IN THE EVENT OF INJURY

This study should not entail any physical or mental risk beyond those described above. We do not expect complications to occur, but if, for any reason, you feel that continuing this study would constitute a hardship for you, we will end your participation in the study.

In the event of a medical emergency while participating in this study or medical treatment required as a result of your participation in this study, you may receive emergency treatment in the facility you are in or a nearby Department of Defense (military) medical facility (hospital or clinic). Treatment/care will be provided even if you are not eligible to receive such care. Care will be continued until the medical doctor treating you decides that you are out of immediate danger. If you are not entitled to care in a military facility, you may be transferred to a private civilian hospital. The attending doctor or member of the hospital staff will go over the transfer decision with you before it happens. The military will bill your health insurance for health care you receive which is not part of the study. You will not be personally billed and you WILL NOT be expected to pay for medical care at our hospitals. If you are required to pay a deductible you may make a claim for reimbursement through the Uniformed Services University Office of General Counsel. In case you need additional care following discharge from the military hospital or clinic, a military health care professional will decide whether your need for care is directly related to being in the study. If your need for care is related to the study, the military may offer you limited health care at its medical facilities. This additional care is not automatic.

If at any time you believe you have suffered an injury or illness as a result of participating in this research project, you should contact the Office of Research at the Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799 at (301) 295-3303. This office can review the matter with you, can provide information about your rights as a subject, and may be able to identify resources available to you. If you believe the government or one of the government's employees (such as a military doctor) has injured you, a claim for damages (money) against the federal government (including the military) may be filed under the Federal Torts Claims Act. Information about judicial avenues of compensation is available from the University's General Counsel at (301) 295-3028.

IF YOU HAVE QUESTIONS OR CONCERNS

If you have questions about this research, you should contact Dr. Mark Ettenhofer, the person in charge of the study. His phone number at USUHS is 301-295-3279. Even in the evening or on weekends, you can leave a message at that number. If you have questions about your rights as a research subject, you should call the Director, Human Research Protections Program at USUHS at (301) 295-9534. He is your representative and has no connection to the researcher conducting this study.

By signing this form you are agreeing that this study has been explained to you, that you understood that explanation, and that you want to take part in this research.

_____ _____

Subject Date of signature

_____ _____

Witness Date of signature

I certify that the research study has been explained to the above individuals, by me or my research staff, and that the individual understands the nature and purpose, the possible risks and benefits associated with taking part in this research study. Any questions that have been raised have been answered.

_____ _____

Investigator Date of signature

Administrative Document 3: Simulator Study Recruitment Flyer

Healthy Volunteers Needed

FOR PARTICIPATION IN EYE TRACKING RESEARCH STUDY

Eligible volunteers will complete computer tasks while an eye tracker records their eye movements.

Volunteers will also take a series of attention, reaction time, memory tasks.

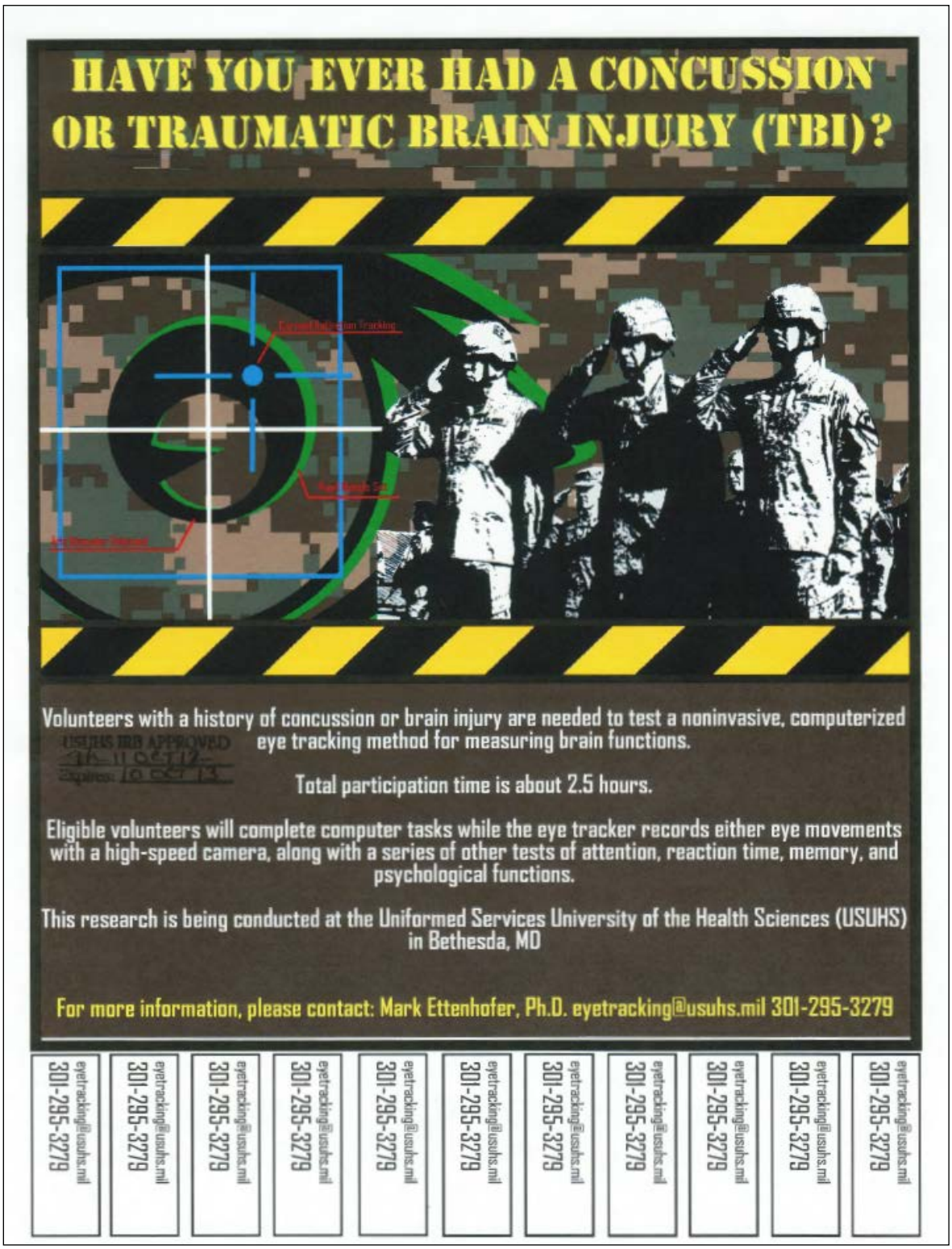
This research is being conducted at the
Uniformed Services University of the Health Sciences (USUHS) in Bethesda, MD

For more information, please contact:
Mark Ettenhofer, Ph.D., at
eyetracking@usuhs.edu
301-295-3279

eyetracking@usuhs.edu 301-295-3279

eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279
eyetracking@usuhs.edu 301-295-3279

Administrative Document 4: Parent Study TBI Cohort Recruitment Flyer



REFERENCES

1. Allen LM, Conder RL, Green P, Cox DR. 1997. *Carb'97: Computerized assessment of response bias. Manual*. Durham, NC: CogniSyst

2. American Psychiatric Association. 1952. *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author

3. American Psychiatric Association. 1968. *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author

4. American Psychiatric Association. 1980. *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author

5. American Psychiatric Association. 1987. *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author

6. American Psychiatric Association. 1994. *Diagnostic and statistical manual of mental disorders*. Washington, DC: Author

7. American Psychiatric Association. 2000. *Diagnostic and statistical manual of mental disorders* Washington, DC: Author

8. Armistead-Jehle P. 2010. Symptom validity test performance in u.S. Veterans referred for evaluation of mild tbi. *Applied Neuropsychology* 17:52-9

9. Armistead-Jehle P, Buican B. 2012. Evaluation context and symptom validity test performances in a u.S. Military sample. *Archives of Clinical Neuropsychology* 27:828-39

10. Armistead-Jehle P, Hansen CL. 2011. Comparison of the repeatable battery for the assessment of neuropsychological status effort index and stand-alone symptom validity tests in a military sample. *Archives of Clinical Neuropsychology* 26:592-601

11. Arnett P. 2013. *Secondary influences on neuropsychological test performance*. New York, NY: Oxford University Press

12. Axelrod BN, Fichtenberg NL, Millis SR, Wertheimer JC. 2006. Detecting incomplete effort with digit span from the wechsler adult intelligence scale-third edition. *The Clinical Neuropsychologist* 20:513-23

13. Axelrod BN, Schutte C. 2010. Analysis of the dementia profile on the medical symptom validity test. *The Clinical Neuropsychologist* 24:873-81

14. Babikian T, Boone KB, Lu P, Arnold G. 2006. Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist* 20:145-59

15. Baldessarini RJ, Finklestein S, Arana GW. 1983. The predictive power of diagnostic tests and the effect of prevalence of illness. *Archives of General Psychiatry* 40:569-73

16. Barnes G. 2008. Cognitive processes involved in smooth pursuit eye movements. *Brain and Cognition* 68:309-26

17. Barrett KE. 2010. *Ganong's review of medical physiology*. McGraw-Hill Medical

18. Barry DM. 2012. Title. Volume:In press

19. Barry DM, Zeckler G, Brewer S, Girard D, Mirochnitchenko D, et al. 2012. Correlates of a novel eye-tracking assessment: Cognition, depression, traumatic stress, combat exposure, and fatigue. In *2012 Military Health System Research Symposium*. Ft. Lauderdale, FL

20. Bauer L, McCaffrey RJ. 2006. Coverage of the test of memory malingering, victoria symptom validity test, and word memory test on the internet: Is test security threatened? *Archives of Clinical Neuropsychology* 21:121-6

21. Bauer L, Yantz CL, Ryan LM, Warden DL, McCaffrey RJ. 2005. An examination of the california verbal learning test ii to detect incomplete effort in a traumatic brain-injury sample. *Applied Neuropsychology* 12:202-7

22. Bazarian JJ, McClung J, Shah MN, Cheng YT, Flesher W, Kraus J. 2005. Mild traumatic brain injury in the united states, 1998--2000. *Brain Injury* 19:85-91

23. Bender SD, Rogers R. 2004. Detection of neurocognitive feigning: Development of a multi-strategy assessment. *Archives of Clinical Neuropsychology* 19:49-60

24. Bernard LC. 1990. Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology* 12:715-28

25. Berry DT, Schipper LJ. 2007. Detection of feigned psychiatric symptoms during forensic neuropsychological examinations. In *Assessment of malingered neuropsychological deficits*, ed. GJ Larrabee:226-63. New York, NY: Oxford University Press. Number of 226-63 pp.

26. Berry DTR, Nelson NW. 2010. Dsm-5 and malingering: A modest proposal. *Psychological Injury and Law* 3:295-303

27. Bianchini KJ, Curtis KL, Greve KW. 2013. Cognitive performance validity assessment in mild traumatic brain injury, physical pain, and posttraumatic stress.

In *Mild traumatic brain injury: Symptom validity assessment and malingering* ed. DA Carone, SS Bush:323-46. New York, NY: Springer Publisher Company. Number of 323-46 pp.

28. Bianchini KJ, Greve KW, Glynn G. 2005. On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal* 5:404-17

29. Bigler ED. 2011. Effort – what is it, how should it be measured? *Journal of the International Neuropsychological Society* 17:751-2

30. Bigler ED. 2012. Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society* 18:632-40

31. Binder LM. 1993. Assessment of malingering after mild head trauma with the portland digit recognition test. *Journal of Clinical and Experimental Neuropsychology* 15:170-82

32. Binder LM, Rohling ML. 1996. Money matters: A meta-analytic review of the effects of financial incentives on recovery after closed-head injury. *The American Journal of Psychiatry* 153:7-10

33. Bleiberg J, Cernich AN, Cameron K, Sun W, Peck K, et al. 2004. Duration of cognitive impairment after sports concussion. *Neurosurgery* 54:1073-78; discussion 8-80

34. Boone KB. 2007. *Assessment of feigned cognitive impairment: A neuropsycholoical perspective*. New York, NY: The Guilford Press

35. Boone KB. 2007. A reconsideration of the slick et al. (1999) criteria for malingered neurocognitive dysfunction. In *Assessment of feigned cognitive impairment: A neuropsycholoical perspective*, ed. KB Boone:29-49. New York, NY: Guilford Press. Number of 29-49 pp.

36. Boone KB. 2009. The need for continuous and comprehensive sampling of effort/response bias during neuropsychological examinations. *The Clinical Neuropsychologist* 23:729-41

37. Boone KB, Lu P, Herzberg D. 2002. *The dot counting test*. Los Angeles, CA: Western Psychological Services

38. Boone KB, Lu P, Sherman D, Palmer B, Back C, et al. 2000. Validation of a new technique to detect malingering of cognitive symptoms: The b test. *Archives of Clinical Neuropsychology* 15:227-41

39. Browndyke JN. 2013. Functional neuroanatomical bases of deceptive behavior and malingering. In *Mild traumatic brain injury: Symptom validity assessment*

and malingering, ed. DA Carone, SS Bush:303-21. New York, NY: Springer Publishing Company. Number of 303-21 pp.

40. Bruhn P, Parsons OA. 1977. Reaction time variability in epileptic and brain-damaged patients. *Cortex*

41. Bush SS, Ruff RM, Troster AI, Barth JT, Koffler SP, et al. 2005. Symptom validity assessment: Practice issues and medical necessity nan policy & planning committee. *Archives of Clinical Neuropsychology* 20:419-26

42. Busse M, Whiteside D. 2012. Detecting suboptimal cognitive effort: Classification accuracy of the conner's continuous performance test-ii, brief test of attention, and trail making test. *The Clinical Neuropsychologist* 26:675-87

43. Butcher JN, Graham JR, Ben-Porath YS, et al. 2001. *Mmpi-2: Manual for administration, scoring, and interpretation*. Minneapolis, MN: University of Minnesota Press

44. Carone DA, Bush SS. 2013. Introduction: Historical perspectives on mild traumatic brain injury, symptom validity assessment, and malingering. In *Mild traumatic brain injury: Symptom validity assessment and malingering*, ed. DA Carone, SS Bush:1-29. New York, NY: Springer Publishing Company. Number of 1-29 pp.

45. Cassidy JD, Carroll LJ, Peloso PM, Borg J, von Holst H, et al. 2004. Incidence, risk factors and prevention of mild traumatic brain injury: Results of the who collaborating centre task force on mild traumatic brain injury. *Journal of Rehabilitation Medicine*:28-60

46. Cavegn D. 1996. Bilateral interactions in saccade programming. A saccade-latency study. *Experimental Brain Research* 109:312-32

47. Centers for Disease Control and Prevention. 2003. *Report to congress on mild traumatic brain injury in the united states: Steps to prevent a serious public health problem*. Atlanta, GA: National Center for Injury Prevention and Control

48. Chafetz M. 2011. Reducing the probability of false positives in malingering detection of social security disability claimants. *The Clinical Neuropsychologist* 25:1239-52

49. Chafetz MD. 2008. Malingering on the social security disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist* 22:529-46

50. Chafetz MD. 2010. Symptom validity issues in the psychological consultative examination for social security disability. *The Clinical Neuropsychologist* 24:1045-63

51. Chafetz MD. 2011. The psychological consultative examination for social security disability. *Psychological Injury and Law*:1-10

52. Chafetz MD, Prentkowski E, Rao A. 2011. To work or not to work: Motivation (not low iq) determines symptom validity test findings. *Archives of Clinical Neuropsychology* 26:306-13

53. Chen H-C, Koh C-L, Hsieh C-L, Hsueh I-P. 2009. Test-re-test reliability of two sustained attention tests in persons with chronic stroke. *Brain Injury* 23:715-22

54. Cicerone KD, Azulay J. 2002. Diagnostic utility of attention measures in postconcussion syndrome. *The Clinical Neuropsychologist* 16:280-9

55. Conners CK. 2004. *Conners' continuous performance test (cpt ii)*. Toronto, Ontario, Canada: Multi-Health Systems, Inc.

56. Coronado VG, McGuire LC, Sarmiento K, Bell J, Lionbarger MR, et al. 2012. Trends in traumatic brain injury in the u.S. And the public health response: 1995-2009. *Journal of Safety Research* 43:299-307

57. Corrigan JD, Selassie AW, Orman JAL. 2010. The epidemiology of traumatic brain injury. *The Journal of Head Trauma Rehabilitation* 25:72

58. Crawford TJ, Higham S, Renvoize T, Patel J, Dale M, et al. 2005. Inhibitory control of saccadic eye movements and cognitive impairment in alzheimer's disease. *Biological Psychiatry* 57:1052-60

59. Curtis KL, Greve KW, Bianchini KJ. 2009. The wechsler adult intelligence scale-iii and malingering in traumatic brain injury: Classification accuracy in known groups. *Assessment* 16:401-14

60. Dallal GE. 2013. *Randomization plans.* www.randomization.com

61. Dearth CS, Berry DTR, Vickery CD, Vagnini VL, Baser RE, et al. 2005. Detection of feigned head injury symptoms on the mmpi-2 in head injured patients and community controls. *Archives of Clinical Neuropsychology* 20:95-110

62. Defense and Veterans Brain Injury Center. 2012. *Dod worldwide numbers for traumatic brain injury*. http://www.dvbic.org/dod-worldwide-numbers-tbi

63. Delis DC, Kramer JH, Kaplan E, Ober BA. 2000. *California verbal learning test-second edition (cvlt-ii).* San Antonio, TX: Psychological Corporation

64. DeLong ER, DeLong DM, Clarke-Pearson DL. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*:837-45

65. Department of Veterans Affairs. 2009. *Compensation and pension (c&p) examinations*. http://www1.va.gov/vhapublications/ViewPublication.asp?pub_ID=2094

66. Department of Veterans Affairs, Department of Defense. 2009. Va/dod clinical practice guideline for management of concussion/mild traumatic brain injury. Washington, DC: Department of Veteran Affairs, Department of Defense

67. Dikmen SS, Heaton RK, Grant I, Temkin NR. 1999. Test–retest reliability and practice effects of expanded halstead–reitan neuropsychological test battery. *Journal of the International Neuropsychological Society* 5:346-56

68. Drew AS, Langan J, Halterman C, Osternig LR, Chou L-S, van Donkelaar P. 2007. Attentional disengagement dysfunction following mtbi assessed with the gap saccade task. *Neuroscience Letters* 417:61-5

69. Duchowski AT. 2007. *Eye tracking methodology: Theory and practice*. London: Springer

70. Ehrenstein WH, Heister G, Cohen R. 1982. Trail making test and visual search. *Archiv fur Psychiatrie und Nervenkrankheiten* 231:333-8

71. Erdal K. 2004. The effects of motivation, coaching, and knowledge of neuropsychology on the simulated malingering of head injury. *Archives of Clinical Neuropsychology* 19:73-88

72. Ettenhofer ML, Abeles N. 2009. The significance of mild traumatic brain injury to cognition and self-reported symptoms in long-term recovery from injury. *Journal of Clinical and Experimental Neuropsychology* 31:363-72

73. Ettenhofer ML, Foley J, Behdin N, Levine AJ, Castellon SA, Hinkin CH. 2010. Reaction time variability in hiv-positive individuals. *Archives of Clinical Neuropsychology* 25:791-8

74. Exton C, Leonard M. 2009. Eye tracking technology: A fresh approach in delirium assessment? *International Review of Psychiatry* 21:8-14

75. Faul F, Erdfelder E, Lang A-G, Buchner A. 2007. G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39:175-91

76. Faul M, Xu L, Wald MM, Coronado VG. 2010. Traumatic brain injury in the united states: Emergency department visits, hospitalizations, and deaths., Centers for Disease Control and Prevention, National Center for Injury Prevention and Control, Atlanta (GA)

77. Feigin VL, Theadom A, Barker-Collo S, Starkey NJ, McPherson K, et al. 2013. Incidence of traumatic brain injury in new zealand: A population-based study. *Lancet Neurol* 12:53-64

78. Fischer B. 1986. The role of attention in the preparation of visually guided eye movements in monkey and man. *Psychological Research* 48:251-7

79. Fischer B. 1987. The preparation of visually guided saccades. *Reviews of Physiology, Biochemistry and Pharmacology* 106:1-35

80. Fischer B, Breitmeyer B. 1987. Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia* 25:73-83

81. Fischer B, Gezeck S, Huber W. 1995. The three-loop model: A neural network for the generation of saccadic reaction times. *Biological Cybernetics* 72:185-96

82. Fox DD. 2011. Symptom validity test failure indicates invalidity of neuropsychological tests. *The Clinical Neuropsychologist* 25:488-95

83. Frederick RI. 1997. *Vip: Validity indicator profile manual*. Minnetonka, MN: National Computer Systems Assessments

84. Frederick RI. 2003. Review of the validity indicator profile. *Journal of Forensic Neuropsychology* 2:125-45

85. Frederick RI, Speed FM. 2007. On the interpretation of below-chance responding in forced-choice tests. *Assessment* 14:3-11

86. Frost RB, Farrer TJ, Primosch M, Hedges DW. 2012. Prevalence of traumatic brain injury in the general adult population: A meta-analysis. *Neuroepidemiology* 40:154-9

87. Furr RM, Bacharach VR. 2008. *Psychometrics : An introduction*. Los Angeles: Sage Publications. xvi, 349 p. pp.

88. Gervais RO, Ben-Porath YS, Wygant DB, Green P. 2007. Development and validation of a response bias scale (rbs) for the mmpi-2. *Assessment* 14:196-208

89. Goebel RA. 1983. Detection of faking on the halstead-reitan neuropsychological test battery. *Journal of Clinical Psychology* 39:731-42

90. Goldberg HE, Back-Madruga C, Boone KB. 2007. The impact of psychiatric disorders on cognitive symptom validity test scores. In *Assessment of feigned cognitive impairment: A neuropsycholoical perspective* ed. KB Boone:281-309. New York, NY: Guilford Press. Number of 281-309 pp.

91. Goldstein ER. 1923. Reaction times and the consciousness of deception. *The American Journal of Psychology*:562-81

92. Gooding DC, Basso MA. 2008. The tell-tale tasks: A review of saccadic research in psychiatric patient populations. *Brain and Cognition* 68:371-90

93. Green P. 2003. *Green's word memory test for windows: User's manual and program*. Edmonton, Alberta, Canada: Author

94. Green P. 2004. *Medical symptom validity test for windows: User's manual and program*. Edmonton, Alberta, Canada: Author

95. Green P. 2007. The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America* 18:43-68, vi

96. Green P. 2007. Spoiled for choice: Making comparisons between forced-chioce effort tests. In *Assessment of feigned cognitive impairment: A neuropsycholoical perspective* ed. KB Boone:50-77. New York, NY: Guilford Press. Number of 50-77 pp.

97. Green P. 2008. *Manual for the nonverbal medical symptom validity test (nv-msvt)*. Edmonton, Canada: Green's Publishing

98. Green P, Allen LM, Astner K. 1996. *The word memory test: A user's guide to the oral and computer-administered forms, us version 1.1*. Durham, NC: Cognisyst

99. Green P, Rohling ML, Lees-Haley PR, Allen LM, 3rd. 2001. Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury* 15:1045-60

100. Greiffenstein MF. 2013. Foreward. In *Mild traumatic brain injury: Symptom validity assessment and malingering* ed. DA Carone, SS Bush:xiii-xiv. New York, NY: Springer Publishing Company. Number of xiii-xiv pp.

101. Greiffenstein MF, Baker WJ, Gola T. 1994. Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment* 6:218

102. Grimes DA, Schulz KF. 2005. Refining clinical diagnosis with likelihood ratios. *The Lancet* 365:1500-5

103. Grote CL, Hook JN. 2007. Forced-choice recognition tests of malingering. In *Assessment of malingered neuropsychological deficits (pp. 131-170). New york, ny: Oxford university press.*, ed. GJ Larrabee:44-79. New York, NY: Oxford University Press. Number of 44-79 pp.

104. Groth-Marnat G. 2009. *Handbook of psychological assessment*. Hoboken, N.J.: John Wiley & Sons, Inc. xiv, 752 p. pp.

105. Guidotti Breting LM, Sweet JJ. 2013. Freestanding cognitive sympom validity tests: Use and selection in mild traumatic brain injury. In *Mild traumatic brain injury: Symptom validity assessment and malingering*, ed. DA Carone, SS

Bush:145-57. New York, NY: Springer Publishing Company. Number of 145-57 pp.

106. Guilmette TJ. 2013. The role of clinical judgment in symptom validity assessment. In *Mild traumatic brain injury: Symptom validity assessment and malingering* ed. DA Carone, SS Bush:31-43. New York, NY: Springer Publishing Company. Number of 31-43 pp.

107. Guyton AC, Hall JE. 2006. *Textbook of medical physiology*. Philadelphia: Elsevier Saunders. xxxv, 1116 p. pp.

108. Guyton AC, Hall JE. 2011. *Textbook of medical physiology*. Philadelphia: Elsevier Saunders. xxxv, 1116 p. pp.

109. Hajian-Tilaki KO, Hanley JA, Joseph L, Collet JP. 1997. A comparison of parametric and nonparametric approaches to roc analysis of quantitative diagnostic tests. *Medical Decision Making* 17:94-102

110. Hanley JA, McNeil BJ. 1982. The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* 743:29-36

111. Hannula DE, Baym CL, Warren DE, Cohen NJ. 2012. The eyes know: Eye movements as a veridical index of memory. *Psychological Science* 23:278-87

112. Hartlage LC. 2012. Clinical detection of malingering. In *Detection of malingering during head injury litigation*, ed. CR Reynolds, AM Horton:333-50. New York, NY: Springer. Number of 333-50 pp.

113. Hartman DE. 2008. The computerized test of information processing (ctip) by tom tombaugh. *Applied Neuropsychology* 15:226-7

114. Heaton RK, Grant I, Matthews CG. 1991. *Comprehensive norms for an expanded halstead-reitan battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources

115. Heaton RK, Smith HH, Lehman RA, Vogt AT. 1978. Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology* 46:892-900

116. Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR. 2009. American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist* 23:1093-129

117. Heitger MH, Jones RD, Macleod AD, Snell DL, Frampton CM, Anderson TJ. 2009. Impaired eye movements in post-concussion syndrome indicate suboptimal brain function beyond the influence of depression, malingering or intellectual ability. *Brain* 132:2850-70

118. Henry GK. 2005. Probable malingering and performance on the test of variables of attention. *The Clinical Neuropsychologist* 19:121-9

119. Henry GK, Enders C. 2007. Probable malingering and performance on the continuous visual memory test. *Applied Neuropsychology* 14:267-74

120. Henry GK, Heilbronner RL, Mittenberg W, Enders C. 2006. The henry-heilbronner index: A 15-item empirically derived mmpi-2 subscale for identifying probable malingering in personal injury litigants and disability claimants. *The Clinical Neuropsychologist* 20:786-97

121. Hiscock M, Hiscock CK. 1989. Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology* 11:967-74

122. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL. 2004. Combat duty in iraq and afghanistan, mental health problems, and barriers to care. *The New England Journal of Medicine* 351:13-22

123. Hoge CW, McGurk D, Thomas JL, Cox AL, Engel CC, Castro CA. 2008. Mild traumatic brain injury in u.S. Soldiers returning from iraq. *The New England Journal of Medicine* 358:453-63

124. Holdnack JA. 2001. *Wechsler test of adult reading: Manual.* San Antonio, TX: Pearson, Inc.

125. Hosmer DW, Lemeshow S. 2000. *Applied logistic regression*. New York, NY: John Wiley and Sons. 160-4 pp.

126. Howe LLS, Anderson AM, Kaufman DAS, Sachs BC, Loring DW. 2007. Characterization of the medical symptom validity test in evaluation of clinically referred memory disorders clinic patients. *Archives of Clinical Neuropsychology* 22:753-61

127. Hsiao JK, Bartko JJ, Potter WZ. 1989. Diagnosing diagnoses. Receiver operating characteristic methods and psychiatry. *Archives of General Psychiatry* 46:664-7

128. Hutton S. 2008. Cognitive control of saccadic eye movements. *Brain and Cognition* 68:327-40

129. Inman TH, Berry DTR. 2002. Cross-validation of indicators of malingering: A comparison of nine neuropsychological tests, four tests of malingering, and behavioral observations. *Archives of Clinical Neuropsychology* 17:1-23

130. Inman TH, Vickery CD, Berry DT, Lamb DG, Edwards CL, Smith GT. 1998. Development and initial validation of a new procedure for evaluating adequacy of effort given during neuropsychological testing: The letter memory test. *Psychological Assessment* 10:128-39

131. Iverson GL. 2005. Outcome from mild traumatic brain injury. *Current Opinion in Psychiatry* 18:301-17

132. Iverson GL, Binder LM. 2000. Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation* 15:829-58

133. Iverson GL, Franzen MD, McCracken LM. 1991. Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior* 15:667-76

134. Iverson GL, Lange RT, Green P, Franzen MD. 2002. Detecting exaggeration and malingering with the trail making test. *The Clinical Neuropsychologist* 16:398-406

135. Ivnik RJ, Smith GE, Cerhan JH, Boeve BF, Tangalos EG, Petersen RC. 2001. Understanding the diagnostic capabilities of cognitive tests. *The Clinical Neuropsychologist* 15:114-24

136. Jackson RL, Rogers R, Sewell KW. 2005. Forensic applications of the miller forensic assessment of symptoms test (mfast): Screening for feigned disorders in competency to stand trial evaluations. *Law and Human Behavior* 29:199-210

137. Jasinski LJ, Berry DTR, Shandera AL, Clark JA. 2011. Use of the wechsler adult intelligence scale digit span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology* 33:300-14

138. Jasinski LJ, Harp JP, Berry DT, Shandera-Ochsner AL, Mason LH, Ranseen JD. 2011. Using symptom validity tests to detect malingered adhd in college students. *The Clinical Neuropsychologist* 25:1415-28

139. Johnson LR. 2003. *Essential medical physiology*. San Diego, CA: Academic Press

140. Johnson R, Barnhardt J, Zhu J. 2003. The deceptive response: Effects of response conflict and strategic monitoring on the late positive component and episodic memory-related brain activity. *Biological Psychology* 64:217-53

141. Kapur N. 1994. The coin-in-the-hand test: A new" bed-side" test for the detection of malingering in patients with suspected memory disorder. *Journal of neurology, neurosurgery, and psychiatry* 57:385

142. Kay T, Harrington D, Adams R, Anderson T, Berrol S. 1993. Definition of mild traumatic brain injury. *Journal of Head Trauma Rehabilitation* 8:86-8

143. Keiser LR, Miller SM. 2010. The impact of organized interests on eligibility determination: The case of veterans' disability compensation. *Journal of Public Administration Research and Theory* 20:505-31

144. Kingstone A, Klein RM. 1993. Visual offsets facilitate saccadic latency: Does predisengagement of visuospatial attention mediate this gap effect? *Journal of Experimental Psychology: Human Perception and Performance* 19:1251

145. Kirkwood MW, Hargrave DD, Kirk JW. 2011. The value of the wisc-iv digit span subtest in detecting noncredible performance during pediatric neuropsychological examinations. *Archives of Clinical Neuropsychology* 26:377-84

146. Kirkwood MW, Kirk JW. 2010. The base rate of suboptimal effort in a pediatric mild tbi sample: Performance on the medical symptom validity test. *The Clinical Neuropsychologist* 24:860-72

147. Kraus MF, Little DM, Donnell AJ, Reilly JL, Simonian N, Sweeney JA. 2007. Oculomotor function in chronic traumatic brain injury. *Cognitive and Behavioral Neurology* 20:170-8

148. Kraus MF, Little DM, Wojtowicz SM, Sweeney JA. 2010. Procedural learning impairments identified via predictive saccades in chronic traumatic brain injury. *Cognitive and Behavioral Neurology* 23:210-7

149. Lange RT, Iverson GL, Brickell TA, Staver T, Pancholi S, et al. 2012. Clinical utility of the conners' continuous performance test-ii to detect poor effort in u.S. Military personnel following traumatic brain injury. *Psychological Assessment*

150. Lange RT, Iverson GL, Brooks BL, Rennison V. 2010. Influence of poor effort on self-reported symptoms and neurocognitive test performance following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology* 32:961-72

151. Lange RT, Pancholi S, Bhagwat A, Anderson-Barnes V, French LM. 2012. Influence of poor effort on neuropsychological test performance in u.S. Military personnel following mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology* 34:453-66

152. Langlois JA, Rutland-Brown W, Wald MM. 2006. The epidemiology and impact of traumatic brain injury: A brief overview. *Journal of Head Trauma Rehabilitation* 21:375-8

153. Larrabee GJ. 2003. Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist* 17:410-25

154. Larrabee GJ. 2007. *Assessment of malingered neuropsychological deficits*. New York, NY: Oxford University Press. xiv, 386 p. pp.

155. Larrabee GJ. 2007. Introduction: Malingering, research designs, and base rates. In *Assessment of malingered neuropsychological deficits*, ed. GJ Larrabee:3-13. New York, NY: Oxford University Press. Number of 3-13 pp.

156. Larrabee GJ. 2012. Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society* 18:625-30

157. Larrabee GJ, Berry DTR. 2007. Diagnostic classification statistics and diagnostic validity of malingering assessment. In *Assessment of malingered neuropsychological deficits* ed. GJ Larrabee:14-26. New York, NY: Oxford University Press. Number of 14-26 pp.

158. Larrabee GJ, Millis SR, Meyers JE. 2009. 40 plus or minus 10, a new magical number: Reply to russell. *The Clinical Neuropsychologist* 23:841-9

159. Larson EB, Kondiles BR, Starr CR, Zollman FS. 2012. Postconcussive complaints, cognition, symptom attribution and effort among veterans. *Journal of the International Neuropsychological Society*:1-8

160. Leark RA, Dixon D, Hoffman T, Huynh D. 2002. Fake bad test response bias effects on the test of variables of attention. *Journal of the International Neuropsychological Society* 17:335-42

161. Leark RA, Dupuy TR, Greenberg L, Kindschi C, Corman CL. 1996. *Tova: Test of variables of attention: Professional manual*. Universal Attention Disorders

162. Lees-Haley PR, English LT, Glenn WJ. 1991. A fake bad scale on the mmpi-2 for personal injury claimants. *Psychological Reports* 68:203-10

163. Lezak MD, Howieson DB, Bigler ED, Tranel D. 2012. *Neuropsychological assessment*. New York, NY: Oxford University Press

164. Lezak MD, Howieson DB, Loring DW. 2004. *Neuropsychological assessment*. pp 1016. New York, NY: Oxford University Press

165. Locke DE, Smigielski JS, Powell MR, Stevens SR. 2008. Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation* 23:273-81

166. LoPiccolo CJ, Goodkin K, Baldewicz TT. 1999. Current issues in the diagnosis and management of malingering. *Annals of Medicine* 31:166-74

167. Maher BA. 1978. A reader's, writer's, and reviewer's guide to assessing research reports in clinical psychology. *Journal of Consulting and Clinical Psychology* 46:835

168. Mandrekar JN. 2010. Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology* 5:1315-6 10.097/JTO.0b013e3181ec173d

169. Marshall P, Schroeder R, O'Brien J, Fischer R, Ries A, et al. 2010. Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. *The Clinical Neuropsychologist* 24:1204-37

170. McCormick CL, Yoash-Gantz RE, McDonald SD, Campbell TC, Tupler LA. 2013. Performance on the green word memory test following operation enduring freedom/operation iraqi freedom-era military service: Test failure is related to evaluation context. *Archives of Clinical Neuropsychology* 28:808-23

171. McCrea M. 2008. *Mild traumatic brain injury and postconcussion syndrome: The new evidence base for diagnosis and treatment*. USA: Oxford University Press

172. McNeil BJ, Hanley JA. 1984. Statistical approaches to the analysis of receiver operating characteristic (roc) curves. *Medical Decision Making* 4:137

173. Meehl PE, Rosen A. 1955. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin* 52:194-216

174. Merriam-Webster. n.d. *Malinger*. http://www.merriam-webster.com/dictionary/malinger

175. Merten T, Green P, Henry M, Blaskewitz N, Brockhaus R. 2005. Analog validation of german-language symptom validity tests and the influence of coaching. *Archives of Clinical Neuropsychology* 20:719-26

176. Merten T, Thies E, Schneider K, Stevens A. 2009. Symptom validity testing in claimants with alleged posttraumatic stress disorder: Comparing the morel emotional numbing test, the structured inventory of malingered symptomatology, and the word memory test. *Psychological Injury and Law* 2:284-93

177. Meyers JE, Millis SR, Volkert K. 2002. A validity index for the mmpi-2. *Archives of Clinical Neuropsychology* 17:157-69

178. Meyers JE, Volbrecht M, Axelrod BN, Reinsch-Boothby L. 2011. Embedded symptom validity tests and overall neuropsychological test performance. *Archives of Clinical Neuropsychology* 26:8-15

179. Miele AS, Gunner JH, Lynch JK, McCaffrey RJ. 2012. Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology* 27:10-22

180. Miller HA. 2001. *Mfast: Miller forensic assessment of symptoms test. Professional manual*. Odessa, FL: Psychological Assessment Resources

181. Millis SR, Volinsky CT. 2001. Assessment of response bias in mild head injury: Beyond malingering tests. *Journal of Clinical and Experimental Neuropsychology* 23:809-28

182. Millon T. 1994. *Manual for the mcmi-iii*. Minneapolis, MN: National Computer Systems

183. Mittenberg W, Patton C, Canyock EM, Condit DC. 2002. Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology* 24:1094-102

184. Morel KR. 1998. Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. *Journal of Personality Assessment* 70:299-314

185. Morel KR, Shepherd BE. 2008. Meta-analysis of the morel emotional numbing test for ptsd: Comment on singh, avasthi, and grover. *German Journal of Psychiatry* 11:128-31

186. Morey LC. 1991. *Personality assessment inventory professional manual*. Odessa, FL: Psychological Assessment Resources

187. Mossman D. 2003. Daubert, cognitive malingering, and test accuracy. *Law and Human Behavior* 27:229-49

188. Mossman D, Wygant DB, Gervais RO. 2012. Estimating the accuracy of neurocognitive effort measures in the absence of a "gold standard". *Psychological Assessment*

189. Munsterberg H. 1914. *Grundzuge der psychotechnik*. Leipzig, Germany: J. A. Barth

190. Müri RM, Nyffeler T. 2008. Neurophysiology and neuroanatomy of reflexive and volitional saccades as revealed by lesion studies with neurological patients and transcranial magnetic stimulation (tms). *Brain and Cognition* 68:284-92

191. Nelson NW, Boone K, Dueck A, Wagener L, Lu P, Grills C. 2003. Relationships between eight measures of suspect effort. *The Clinical Neuropsychologist* 17:263-72

192. Nelson NW, Doane BM. 2013. Research and symptom validity assessment in mild traumatic brain injury cases. In *Mild traumatic brain injury: Symptom validity assessment and malingering*, ed. DA Carone, SS Bush:119-44. New York, NY: Springer Publishing Company. Number of 119-44 pp.

193. Nelson NW, Hoelzle JB, McGuire KA, Ferrier-Auerbach AG, Charlesworth MJ, Sponheim SR. 2010. Evaluation context impacts neuropsychological performance of oef/oif veterans with reported combat-related concussion. *Archives of Clinical Neuropsychology* 25:713-23

194. Nigg JT, Stavro G, Ettenhofer M, Hambrick DZ, Miller T, Henderson JM. 2005. Executive functions and adhd in adults: Evidence for selective effects on adhd symptom domains. *Journal of Abnormal Psychology* 114:706

195. O'Bryant SE, Hilsabeck RC, Fisher JM, McCaffrey RJ. 2003. Utility of the trail making test in the assessment of malingering in a sample of mild traumatic brain injury litigants. *The Clinical Neuropsychologist* 17:69-74

196. Olk B, Kingstone A. 2003. Why are antisaccades slower than prosaccades? A novel finding using a new paradigm. *Neuroreport* 14:151

197. Ono M, Ownsworth T, Walters B. 2011. Preliminary investigation of misconceptions and expectations of the effects of traumatic brain injury and symptom reporting. *Brain Injury* 25:237-49

198. Ord JS, Boettcher AC, Greve KW, Bianchini KJ. 2010. Detection of malingering in mild traumatic brain injury with the conners' continuous performance test-ii. *Journal of Clinical and Experimental Neuropsychology* 32:380-7

199. Pankratz L. 1979. Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology* 47:409-10

200. Pankratz L, Fausti SA, Peed S. 1975. A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology* 43:421-2

201. Parker J, Barry DM, Reinhardt L, Mirochnitchenko D, Zeckler G, et al. 2013. Convergent validity of bethesda eye & attention measure (beam) saccadic inhibition errors with measures of executive functioning: A preliminary study. In *41st Annual Meeting of the International Neuropsychological Society*. Honolulu, HI

202. Payne JL. 1974. Fishing expedition probability: The statistics of post hoc hypothesizing. *Polity*:130-8

203. Pierrot-Deseilligny C, Milea D, Müri RM. 2004. Eye movement control by the cerebral cortex. *Current Opinion in Neurology* 17:17-25

204. Posner MI, Rothbart MK. 2007. Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology* 58:1-23

205. Powell MR, Locke DE, Smigielski JS, McCrea M. 2011. Estimating the diagnostic value of the trail making test for suboptimal effort in acquired brain injury rehabilitation patients. *The Clinical Neuropsychologist* 25:108-18

206. Powers R. n.d. *Punitive articles of the ucmj: Article 115—malingering*. http://usmilitary.about.com/od/punitivearticles/a/mcm115.htm

207. Raszkiewicz S. 2010. *The clinical utility of the conners' continuous performance test in the evaluation of youth with conduct disorder*. (Doctoral dissertation)

208. Razali NM, Wah YB. 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics* 2:21-33

209. Recovering Warrior Task Force. 2012. *Reference handbook of key topics and terms*. http://dtf.defense.gov/rwtf/rwtfref5.pdf

210. Reese CS, Suhr JA, Riddle TL. 2012. Exploration of malingering indices in the wechsler adult intelligence scale-digit span subtest. *Archives of Clinical Neuropsychology* 27:176-81

211. Reicker LI. 2008. The ability of reaction time tests to detect simulation: An investigation of contextual effects and criterion scores. *Archives of Clinical Neuropsychology* 23:419-31

212. Reitan RM, Wolfson D. 1985. *The halstead-reitan neuropsychological test battery. Theory and clinical interpretation*. Tucson, AZ: Neuropsychology Press

213. Rey A. 1941. L'examen psychologique dans les cas d'encephalopathie traumatigue [the psychological exam in cases of traumatic encephalopathy]. *Archives de Psychologie* 28:286-340

214. Rey A. 1964. *L'examen clinique en psychologie [the clinical exam in psychology]*. Paris, France: Press Universitaire de France

215. Reynolds CR, Horton AM. 2012. *Detection of malingering during head injury litigation*. New York: Springer. xii, 381 p. pp.

216. Reznek L. 2005. The rey 15-item memory test for malingering: A meta-analysis. *Brain Injury* 19:539-43

217. Rogers R. 1997. Researching dissimulation. In *Clinical assessment of malingering and deception*, ed. R Rogers:398-426. New York, NY: Guilford Press. Number of 398-426 pp.

218. Rogers R. 2008. Researching response styles. In *Clinical assessment of malingering and deception*, ed. R Rogers:411-34. New York, NY: Guilford Press. Number of 411-34 pp.

219. Rogers R, Bagby RM, Dickens SE. 1992. *Structured interview of reported symptoms (sirs): Professional manual*. Odessa, FL: Psychological Assessment Resources

220. Rogers R, Harrell EH, Liff CD. 1993. Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review* 13:255-74

221. Rogers R, Salekin RT, Sewell KW, Goldstein A, Leonard K. 1998. A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. *Law and Human Behavior* 22:353-67

222. Rogers R, Sewell KW, Goldstein AM. 1994. Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior* 18:543-52

223. Rohling ML, Boone KB. 2007. Future directions in effort assessment. In *Assessment of feigned cognitive impairment: A neuropsycholoical perspective*, ed. KB Boone:453-69. New York, NY: Guilford Press. Number of 453-69 pp.

224. Rosenfeld B, Sands SA, Van Gorp WG. 2000. Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology* 15:349-59

225. Ruffolo LF, Guilmette TJ, Willis GW. 2000. Comparison of time and error rates on the trail making test among patients with head injuries, experimental malingerers, patients with suspect effort on testing, and normal controls. *The Clinical Neuropsychologist* 14:223-30

226. Russo AC. 2012. Symptom validity test performance and consistency of self-reported memory functioning of operation enduring freedom/operation iraqi freedom veterans with positive veteran health administration comprehensive traumatic brain injury evaluations. *Archives of Clinical Neuropsychology*

227. Samuel RZ, Mittenberg W. 2005. Focus points. *Primary psychiatry* 12:60-8

228. Sarter M, Gehring WJ, Kozak R. 2006. More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews* 51:145-60

229. Schagen S, Schmand B, de Sterke S, Lindeboom J. 1997. Amsterdam short-term memory test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology* 19:43-51

230. Schatz P, Browndyke J. 2002. Applications of computer-based neuropsychological assessment. *The Journal of Head Trauma Rehabilitation* 17:395

231. Schmidt M. 1996. *Rey auditory verbal learning test: A handbook*. Los Angeles, CA: Western Psychological Services

232. Schretlen D, Brandt J, Krafft L, Van Gorp W. 1991. Some caveats in using the rey 15-item memory test to detect malingered amnesia. *Psychological Assessment* 3:667

233. Schroeder RW, Twumasi-Ankrah P, Baade LE, Marshall PS. 2012. Reliable digit span: A systematic review and cross-validation study. *Assessment* 19:21-30

234. Schutte C, Axelrod BN. 2013. Use of embedded cognitive symptom validity measures in mild traumatic brain injury cases. In *Mild traumatic brain injury: Symptom validity assessment and malingering* ed. DA Carone, SS Bush:159-81. New York, NY: Springer Publishing Company. Number of 159-81 pp.

235. Schwartz BD, O'Brien BA, Evans WJ, McDermott BE, Sautter FJ, Jr., Winstead DK. 1995. Abnormal saccadic eye movements associated with positive family history schizophrenics. *Biological Psychiatry* 38:487-91

236. Seashore CE, Lewis D, Saetveit DL. 1960. *Seashore measures of musical talents* New York, NY: Psychological Corporation

237. Selassie AW, Zaloshnja E, Langlois JA, Miller T, Jones P, Steiner C. 2008. Incidence of long-term disability following traumatic brain injury hospitalization, united states, 2003. *The Journal of Head Trauma Rehabilitation* 23:123-31

238. Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591-611

239. Sharland MJ, Gfeller JD. 2007. A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology* 22:213-23

240. Sharpe JA. 2008. Neurophysiology and neuroanatomy of smooth pursuit: Lesion studies. *Brain and Cognition* 68:241-54

241. Shum DH, O'Gorman JG, Alpar A. 2004. Effects of incentive and preparation time on performance and classification accuracy of standard and malingering-specific memory tests. *Archives of Clinical Neuropsychology* 19:817-23

242. Silver JM. 2012. Effort, exaggeration and malingering after concussion. *Journal of Neurology, Neurosurgery and Psychiatry* 83:836-41

243. Slick DJ, Hopp G, Strauss E, Thompson G. 1997. *Vsvt, victoria symptom validity test: Version 1.0, professional manual*. Psychological Assessment Resources Odessa, FL

244. Slick DJ, Sherman EM, Iverson GL. 1999. Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist* 13:545-61

245. Slick DJ, Sherman EMS. 2012. Differential diagnosis of malingering and related clinical presentations. In *Pediatric forensic neuropsychology*, ed. EMS Sherman, BL Brooks:113-35. New York, NY: Oxford University Press. Number of 113-35 pp.

246. Slick DJ, Sherman EMS. 2013. Differential diagnosis of malingering. In *Mild traumatic brain injury: Symptom validity assessment and malingering*, ed. DA Carone, SS Bush:57-72. New York, NY: Springer Publishing Company. Number of 57-72 pp.

247. Slick DJ, Tan JE, Strauss EH, Hultsch DF. 2004. Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology* 19:465-73

248. Sollman MJ, Berry DT. 2011. Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology* 26:774-89

249. Stevens A, Merten T. 2010. Psychomotor retardation: Authentic or malingered? A comparative study of subjects with and without traumatic brain injury and experimental simulators. *German Journal of Psychiatry*

250. Strauss E, Hultsch DF, Hunter M, Slick DJ, Patry B, Levy-Bencheton J. 2000. Using intraindividual variability to detect malingering in cognitive performance. *The Clinical Neuropsychologist* 14:420-32

251. Strauss E, Slick DJ, Levy-Bencheton J, Hunter M, MacDonald SW, Hultsch DF. 2002. Intraindividual variability as an indicator of malingering in head injury. *Archives of Clinical Neuropsychology* 17:423-44

252. Suhr JA, Barrash J. 2007. Performance on standard attention, memory, and psychomotor speed tasks as indicators of malingering. In *Assessment of malingered neuropsychological deficits* ed. GJ Larrabee:131-70. New York, NY: Oxford University Press. Number of 131-70 pp.

253. Suhr JA, Sullivan BK, Rodriguez JL. 2011. The relationship of noncredible performance to continuous performance test scores in adults referred for attention-deficit/hyperactivity disorder evaluation. *Archives of Clinical Neuropsychology* 26:1-7

254. Sweet JJ, King JH, Malina AC, Bergman MA, Simmons A. 2002. Documenting the prominence of forensic neuropsychology at national meetings and in relevant professional journals from 1990 to 2000. *The Clinical Neuropsychologist* 16:481-94

255. Sweet JJ, Nelson NW. 2007. Validity indicators within executive function measures: Use and limits in detection of malingering. In *Assessment of feigned cognitive impairment: A neuropsycholoical perspective* ed. KB Boone:152-77. New York, NY: Guilford Press. Number of 152-77 pp.

256. Swets JA. 1973. The relative operating characteristic in psychology. *Science* 182:990-1000

257. Tan JE, Slick DJ, Strauss E, Hultsch DF. 2002. How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist* 16:495-505

258. Tanielian T, Jaycox L. 2008. *Invisible wounds of war: Psychological and cognitive injuries, their consequences, and services to assist recovery*. Arlington, VA: Rand Corp. 453 pp.

259. Taylor BC, Hagel EM, Carlson KF, Cifu DX, Cutting A, et al. 2012. Prevalence and costs of co-occurring traumatic brain injury with and without psychiatric disturbance and pain among afghanistan and iraq war veteran va users. *Medical Care* 50:342-6

260. Terrio H, Brenner LA, Ivins BJ, Cho JM, Helmick K, et al. 2009. Traumatic brain injury screening: Preliminary findings in a us army brigade combat team. *Journal of Head Trauma Rehabilitation* 24:14-23

261. Thompson GB. 2003. The victoria symptom validity test: An enhanced test of symptom validity. *Journal of Forensic Neuropsychology* 2:43-67

262. Thurman D, Alverson C, Dunn K, Guerrero J, Sniezek J. 1999. Traumatic brain injury in the united states: A public health perspective. *Journal of Head Trauma Rehabilitation* 14:602-15

263. Tombaugh TN. 1996. *Test of memory malingering (tomm)*. North Tonawanda, NY: Multi Health Systems

264. Tombaugh TN, Rees L, Stormer P, Harrison AG, Smith A. 2007. The effects of mild and severe traumatic brain injury on speed of information processing as measured by the computerized tests of information processing (ctip). *Archives of Clinical Neuropsychology* 22:25-36

265. Trueblood W, Schmidt M. 1993. Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology* 15:578-90

266. Tsanadis J, Montoya E, Hanks RA, Millis SR, Fichtenberg NL, Axelrod BN. 2008. Brain injury severity, litigation status, and self-report of postconcussive symptoms. *The Clinical Neuropsychologist* 22:1080-92

267. Tsirka V, Simos P, Vakis A, Vourkas M, Arzoglou V, et al. 2010. Material-specific difficulties in episodic memory tasks in mild traumatic brain injury. *International Journal of Neuroscience* 120:184-91

268. Van De Graaff KM, Fox SI, Thouin LG. 1999. *Concepts of human anatomy & physiology*. New York, NY: McGraw-Hill College

269. Van Dyke SA, Millis SR, Axelrod BN, Hanks RA. 2013. Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist* 27:1234-46

270. van Stockum S, MacAskill M, Anderson T, Dalrymple-Alford J. 2008. Don't look now or look away: Two sources of saccadic disinhibition in parkinson's disease? *Neuropsychologia* 46:3108-15

271. Vendemia JMC, Buzan RF, Simon-Dack SL. 2005. Reaction time of motor responses in two-stimulus paradigms involving deception and congruity with varying levels of difficulty. *Behavioural neurology* 16:25

272. Vickery CD, Berry DT, Inman TH, Harris MJ, Orey SA. 2001. Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology* 16:45-73

273. Walker R, Kentridge RW, Findlay JM. 1995. Independent contributions of the orienting of attention, fixation offset and bilateral stimulation on human saccadic latencies. *Experimental Brain Research* 103:294-310

274. Warrington EK. 1984. *Recognition memory test*. Los Angeles, CA: Western Psychological Services

275. Weber H, Fischer B. 1995. Gap duration and location of attention focus modulate the occurrence of left/right asymmetries in the saccadic reaction times of human subjects. *Vision Research* 35:987-98

276. Wechsler D. 1997. *Wais-iii: Administration and scoring manual*. San Antonio: Psychological Corp.

277. Wechsler D. 2003. *Wechsler intelligence scale for children-4th edition (wisc-iv): Administration and scoring manual*. San Antonio, TX: Psychological Corporation

278. Wechsler D. 2008. *Wechsler adult intelligence scale-4th edition (wais-iv)*. San Antonio, TX: Harcourt Assessment

279. Weinborn M, Woods SP, Nulsen C, Leighton A. 2012. The effects of coaching on the verbal and nonverbal medical symptom validity tests. *The Clinical Neuropsychologist* 26:832-49

280. Whitney KA, Shepard PH, Williams AL, Davis JJ, Adams KM. 2009. The medical symptom validity test in the evaluation of operation iraqi freedom/operation enduring freedom soldiers: A preliminary study. *Archives of Clinical Neuropsychology* 24:145-52

281. Widows MR, Smith GP. 2005. *Structured inventory of malingered symptomatology (sims): Professional manual*. Lutz, FL: Psychological Assessment Resources

282. Willison J, Tombaugh TN. 2006. Detecting simulation of attention deficits using reaction time tests. *Archives of Clinical Neuropsychology* 21:41-52

283. Young JC, Roper BL, Sawyer RJ. 2011. Symptom validity test use and base rates of failure in the va healthcare system. In *9th annual meeting for the American Academy of Clinical Neuropsychology*. Washington, DC

284. Young JC, Sawyer RJ, Roper BL, Baughman BC. 2012. Expansion and re-examination of digit span effort indices on the wais-iv. *The Clinical Neuropsychologist* 26:147-59

285. Zaloshnja E, Miller T, Langlois JA, Selassie AW. 2008. Prevalence of long-term disability from traumatic brain injury in the civilian population of the united states, 2005. *Journal of Head Trauma Rehabilitation* 23:394-400

286. Zasler ND, Martelli MF. 2003. Mild traumatic brain injury: Impairment and disability assessment caveats. *Neuropsychological Rehabilitation* 1:31-41